



**Politecnico
di Torino**

Politecnico di Torino

Corso di Laurea Magistrale in
Ingegneria Gestionale
A.a. 2021/2022
Sessione di Laurea: Ottobre 2022

Mercati digitali e Online Privacy

Analisi delle differenze riscontrate tra il traffico dei siti Multibrand con
quelli Monobrand prima e dopo l'entrata in vigore del GDPR

Relatori:

Chiar.mo Prof. *Carlo Cambini*

Corelatore:

Chiar.mo Prof. *Lorien Sabatino*

Candidati:

Giovanni Gilardi

Matteo Fosforo



**Politecnico
di Torino**



**Politecnico
di Torino**



Politecnico
di Torino

Ringraziamenti

Per questa tesi vorrei ringraziare i professori Carlo Cambini e Lorien Sabatino per avermi supportato lungo tutte le fasi del percorso con consigli e aiuti fondamentali per la riuscita del lavoro. È stato davvero un privilegio poter lavorare insieme a loro.

Vorrei ringraziare il mio compagno di avventure di questa tesi, Matteo, non ché fedele collega di studi che mi ha affiancato esame dopo esame in questo percorso di Laurea Magistrale con il suo sorriso e forza di volontà, fonti di ispirazione per qualsiasi persona lo affiancherà nel corso degli anni.

Vorrei inoltre ringraziare la mia famiglia: Marty ed Eli le mie piccoline (che tanto piccole ormai non sono), le mie sorelle che ogni giorno illuminano le giornate con i loro sorrisi e la loro energia. Siete due forze della natura, farete ciò che vorrete nella vita, supererete ogni difficoltà e ogni ostacolo con le vostre immense capacità che rendono ogni singola azione di estrema facilità. Sarò sempre lì dietro di voi, a proteggervi, con la mia folta barba pronto a festeggiarvi e a supportarvi in ogni successo e traguardo.

Ringrazio mio papà: sappiamo tutti cosa abbiamo passato in questi ultimi mesi pieni di dolore e magari qualche incomprensione. So solo che sarò sempre il tuo Giò che ti guarderà con infinita ammirazione quando risponderai all'ennesima domanda assurda di Trivial, all'ennesima didascalia del Bartezzaghi azzeccata con naturalezza disarmante o quando racconterai il tuo prossimo aneddoto sportivo che ascolterò come un bambino ascolta il proprio insegnante il primo giorno di scuola. Ti voglio bene e te ne vorrò per sempre.

Ringrazio la mia Gege: con la tua ambizione e determinazione sei la mia fonte di ispirazione, con il tuo sorriso e la tua sincerità illumini ogni mia giornata. Sei stata la mia ancora in questi mesi bui e non smetterò mai di ringraziarti per il tuo essere così presente, comprensiva e sensibile. Hai condiviso con me dolori e sofferenza, ora è tempo di condividere insieme i successi e i traguardi. Ti amo.

Ringrazio i miei nonni per essere sempre presenti nella mia vita come fonti di sapienza ed esperienza. Vi voglio un bene immenso e sarò sempre il vostro piccolo Giò.



**Politecnico
di Torino**

Ringrazio Matte, Tommy e Dudy che sono compagni di vita e di calcio: mi avete fatto ridere, divertire e mi avete fatto capire davvero cosa sia il senso dell'amicizia con la vostra semplicità e capacità di ascolto. Vi voglio bene e ci sarò sempre per voi.

Infine, arrivi tu mamma: ogni singola parola di questa tesi è stata scritta pensando a te. Così come ogni cosa che faccio da quel maledetto 4 gennaio. Non trovo nemmeno le parole per descrivere cosa sei stata per me, perché forse non le hanno ancora inventate. Sai che non sono una persona di tante parole, sono riservato nei confronti delle mie emozioni e preferisco far parlare i fatti piuttosto che i sentimenti.

Questa tesi la dedico a te, come dedico a te ogni mio successo, ogni mio traguardo che la vita mi riserverà: se mi hai detto di non essere mai sazio di conoscenza e di successi, sappi che non lo sarò mai perché ogni mia vittoria avrà il tuo nome. Con amore, Giò.



Abstract

La seguente tesi si pone l'obiettivo di valutare l'impatto che l'introduzione del GDPR (General Data Protection Regulation) ha avuto sul traffico dei siti Multibrand e Monobrand, in ambito E-commerce.

Lo studio è stato condotto a partire da un'approfondita literature review suddivisa in quattro tematiche: mercato dei dati, GDPR, governance and regulation e privacy.

Questa prima parte ha permesso di comprendere quali fossero i trend della letteratura scientifica in questi ultimi anni e quali fossero i principali aspetti evidenziati nelle varie branche scientifiche e lo stato attuale dell'arte.

In seguito, si sono analizzati gli articoli prodotti dal gruppo di ricerca del Prof. Carlo Cambini insieme al Prof. Lorien Sabatino attinenti agli ambiti di studio di questa tesi, comprendendo quali fossero le conclusioni e i risultati raggiunti, in ottica di allineare lo studio proposto con i risultati già ottenuti.

Nel terzo capitolo invece si propone lo studio che è stato effettuato sul dataset disponibile relativo al traffico online a partire dall'introduzione al dataset attraverso l'utilizzo di statistiche descrittive: si è condotto uno studio per la suddivisione dei siti in Monobrand e Multibrand, andandoli a classificare sia da un punto di vista della dimensione che da un punto di vista geografico.

Successivamente si è passati allo studio dei trend temporali delle variabili disponibili all'interno del dataset, arrivando a risultati significativi per quanto riguarda il confronto tra siti Multibrand e Monobrand, sia da un punto di vista settoriale che geografico, in quanto si sono effettuate differenti comparazione sulla base sia del settore merceologico di appartenenza del sito (Monobrand vs Multibrand) che in base al Paese di provenienza della ricerca (Europa vs Stati Uniti), avendo suddiviso i domini secondo diversi criteri.

I confronti che sono stati effettuati sono:

- Europa vs Stati Uniti per siti Multibrand
- Europa vs Stati Uniti per siti Monobrand
- Multibrand vs Monobrand in Europa
- Multibrand vs Monobrand (Europa + Stati Uniti)

L'ultimo capitolo invece introdurrà l'analisi di regressione effettuata attraverso l'utilizzo del software STATA che ha permesso di confermare i risultati analizzati in precedenza da un punto di vista di significatività statistica, permettendo di concludere lo studio raggiungendo risultati significativi e soddisfacenti.



Sommario

Indice figure	9
Indice tabelle.....	11
1. Introduzione: dati digitali e privacy online nell'era della connessione	12
2. Literature Review.....	14
2.1 Approccio metodologico	14
2.2 Literature review: data market	16
2.3 Literature review: GDPR.....	22
2.3.1 Introduzione al GDPR.....	22
2.3.2 Approccio metodologico per la literature review	23
2.3 Literature review: Governance and Regulation	28
2.4 Literature review: Privacy.....	33
3. Studio di modelli applicati al mercato digitale	37
3.1 Online Privacy and Market Structure: Theory and Evidence (Sabatino L., Sapi G., 2021) 38	
3.1.1 Introduzione	38
3.1.2 Dataset.....	38
3.1.3 Risultati.....	39
3.2 The Impact of Privacy Regulation on Web Traffic Evidence From the GDPR (Sabatino L., Congiu R., Sapi G., 2022)	41
3.2.1 Introduzione	41
3.2.2 Dataset.....	43
3.2.3 Risultati.....	45
3.3 Conclusione.....	48
4. Introduzione al dataset di riferimento e statistiche descrittive sulla base della domanda di ricerca.....	49
4.1 Domanda di ricerca.....	49
4.2 Introduzione al dataset e statistiche descrittive.....	51
4.3 Analisi dei trend temporali del traffico online.....	65
4.4 Conclusioni sulla base dell'analisi dei trend temporali delle variabili individuate.	80
5. Analisi del modello di regressione	82
6. Conclusioni e riflessioni finali.....	90



**Politecnico
di Torino**

<i>Appendice A</i>	92
<i>Bibliography</i>	95



Indice figure

Figura 1: approccio metodologico della literature review.....	15
Figura 2: approccio metodologico alla literature review sul mercato dei dati digitali ..	16
Figura 3: Data Market Co-occurrence Map	17
Figura 4: Mappa riassuntiva delle principali iniziative del GDPR (Carvalho, et al.).....	23
Figura 5: approccio metodologico alla literature review sul GDPR.....	24
Figura 6: GDPR Co-occurrence Map	25
Figura 7: approccio metodologico alla literature review sulla Governance and Regulation.....	28
Figura 8: Co-occurrence Map Governance and Regulation.....	29
Figura 9: approccio metodologico alla literature review sulla Privacy	33
Figura 10: Co-occurrence Map Privacy	34
Figura 11: Trend di variazione revenues online market.....	40
Figura 12: Cumulata delle sanzioni dovute al GDPR	42
Figura 13: Distribuzione del traffico online per canale	45
Figura 14: Variazione del traffico web.....	46
Figura 15: Procedura metodologica applicata per lo studio proposto	50
Figura 16: Procedura per l'individuazione della suddivisione dei siti sulla base della loro dimensione in termine di n° di visite totali	55
Figura 17: Dimensione siti Multibrand	57
Figura 18: Percentuale di suddivisione in base alla dimensione (Multibrand)	57
Figura 19: Dimensioni siti Monobrand	58
Figura 20: Percentuale di suddivisione in base alla dimensione (Monobrand)	58
Figura 21: Visite per sito (Multibrand)	59
Figura 22: Visite per sito (Monobrand)	59
Figura 23: Procedura utilizzate per la segmentazione geografica dei siti.....	60
Figura 24: Suddivisione geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento	61
Figura 25: Suddivisione percentuale geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento.....	62
Figura 26: Suddivisione geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento per la categoria Multibrand.....	63
Figura 27: Suddivisione percentuale geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento per i siti Multibrand	63
Figura 28: Suddivisione geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento per la categoria Monobrand.....	64
Figura 29: Suddivisione percentuale geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento per i siti Monobrand	64
Figura 30: Procedura utilizzata per la valutazione dei trend temporali.....	66
Figura 31: Traffico medio a confronto per siti Multibrand (EU vs US).....	67



Figura 32: Durata media della visita siti Multibrand (EU vs US).....	68
Figura 33: N° medio di pagine visitate di siti Multibrand (EU vs US)	69
Figura 34: Bounce rate medio per i siti Multibrand (EU vs US).....	70
Figura 35: Traffico medio per i siti Monobrand (EU vs US).....	71
Figura 36: Durata media della visita per i siti Monobrand (EU vs US)	72
Figura 37: N° medio di pagine visitate per i siti Monobrand (EU vs US)	73
Figura 38: Bounce rate medio per i siti Monobrand (EU vs US).....	74
Figura 39: Traffico Medio in EU (Multibrand vs Monobrand).....	75
Figura 40: Durata media della visita in EU (Multibrand vs Monobrand)	75
Figura 41: N° medio di pagine visitate in EU (Multibrand vs Monobrand)	76
Figura 42: Bounce rate medio in EU (Multibrand vs Monobrand).....	77
Figura 43: Confronto tra i differenziali US e EU delle variabili individuate tra siti Multibrand e siti Monobrand	78



**Politecnico
di Torino**

Indice tabelle

Tabella 1: Market Share, Devices & App Stores	19
Tabella 2: Market Share, Search Industry	19
Tabella 3: Market Share, E-commerce	20
Tabella 4: Market Share, Social Media	20
Tabella 5: Suddivisione siti Multibrand vs Monobrand.....	53
Tabella 6: Suddivisione percentuale siti Multibrand vs Monobrand	54
Tabella 7: Elenco siti Multibrand individuati	93
Tabella 8: Elenco siti Monobrand individuati	94



1. Introduzione: dati digitali e privacy online nell'era della connessione

Oggi il mondo è popolato da dati, l'essere umano è circondato e influenzato da informazioni che, volente o nolente, riceve da fonti infinite senza aver alcun potere per limitare questo flusso che si imbatte sul suo pensiero e sul suo modo di vivere.

I dati influenzano le decisioni dei cittadini in funzione delle loro preferenze o a seconda delle loro attitudini e comportamenti, andando a violare, sotto certi aspetti, l'incolumità della privacy personale.

Per fornire un framework iniziale al lettore, si riportano alcuni valori relativi al traffico dei dati e informazioni sui maggiori portali online (Nicita A., 2020): ogni minuto su Facebook, una delle fonti principali di dati e informazioni al mondo, vengono creati quasi 3,5 milioni di post, sotto i quali si possono trovare mezzo milione di commenti e quasi 300.000 stati aggiornati.

Un altro esempio è Twitter: sul portale dedicato alla divulgazione di notizie e aggiornamenti da tutto il mondo vengono scambiati, sempre ogni 60 secondi, quasi mezzo milione di tweet.

Sull'applicazione di messaggistica più diffusa nel vecchio continente (Whatsapp) vengono scambiati quasi 30 milioni di messaggi al minuto, mentre su motore di ricerca di Google vengono effettuate quasi 4 milioni di ricerche, sempre considerando il minuto come unità temporale di riferimento.

Quindi risulta essere chiaro e talvolta angosciante che gli utenti e i cittadini sono invasi, direttamente o indirettamente, da una mole di notizie e dati immensa che possono, da una parte, portare il diretto interessato a conoscenza di fatti e notizie, dall'altro lato possono veicolare alcune scelte e decisioni inopportune.

Il risultato finale è che i comportamenti e le decisioni dell'essere umano possono essere facilmente influenzabili da informazioni e notizie che pervengono da fonti differenti, senza che il soggetto stesso se ne renda conto.

La domanda ora che ci si deve porre è la seguente: in un mondo così connesso, denso di informazioni, come si può tutelare la propria privacy?

Ovviamente la risposta è molto complessa, per tale motivo in questa tesi si forniranno diversi spunti fondati su una base quantitativa analitica di diversi studi effettuati sul traffico dei dati e la privacy online.

Si è capito quindi che il dato e la privacy sono due facce della stessa medaglia e sono sicuramente una delle tematiche più rilevanti nei giorni nostri perché ogni azione può



essere facilmente tracciata, archiviata ed utilizzata al fine di monitorare quelle che sono le abitudini e le preferenze dell'individuo in maniera da veicolarne le scelte e le decisioni. Fino ad ora abbiamo parlato solo di un attore, l'essere umano, ma è evidente che non possa essere l'unico protagonista di questa complessa situazione: infatti le imprese traggono enormi benefici dalla divulgazione e condivisione di dati ed informazioni, ma soprattutto dalla corrispondente analisi che permette loro di individuare possibili trend di mercato presenti e futuri, verso cui indirizzare le prospettive dei loro business.

Le imprese hanno quindi compreso l'enorme potenziale dei dati e di tutto il mercato che si è creato relativo alla loro condivisione: oggi il dato è il vero oggetto dei mercati perché grazie a questo le imprese impostano la loro strategia di business, le campagne pubblicitarie e le caratteristiche del prodotto finale.

Bisogna comprendere quindi, come l'individuo possa individuare autonomamente quali siano le informazioni e i dati personali sensibili e quelli invece che risultano irrilevanti ai fini della tutela della privacy personale: ecco quindi che si sono andate ad identificare quelle che possono essere le tematiche rilevanti ai fini di individuare i fattori di maggiore impatto sulle scelte e le decisioni dell'utente in privacy (Acquisti A., 2015).

La prima tematica è relativa all'incertezza delle persone sull'origine e natura dei trade-offs in materia di privacy e sulle loro preferenze; il secondo tema è relativo l'inclinazione da parte dell'individuo ad essere affine alla privacy in funzione del contesto in cui si trova (alcune volte può reputarla un aspetto rilevante, altre meno); il terzo tema è la malleabilità delle preferenze in ambito privacy, cioè l'influenza da parte dei maggiori esperti rispetto alle preferenze dei soggetti meno esperti.

Un secondo tema che si collega direttamente con la conoscenza e l'importanza che l'individuo attesta alla privacy è la determinazione da parte dell'individuo stesso di quelli che sono i confini tra la sfera pubblica e quella privata: è importante che il soggetto protegga le informazioni private e scelga quali possano essere i dati da condividere in rete.

Questo confine, tra sfera pubblica e privata, è estremamente sottile in quanto è strettamente correlato alla rilevanza dell'asimmetria informativa che c'è tra l'utente e gli attori che utilizzano il dato: la raccolta e l'utilizzo dei dati sono diventati pressoché impercettibili all'utente e questo comporta una conoscenza limitata su quelle che in realtà sono le informazioni che l'utente sta condividendo in quel momento in rete.

In conclusione, è opportuno definire il dato come oggetto dei desideri delle economie digitali ma è altrettanto rilevante sottolineare come il contenuto e la tutela del dato stesso debbano essere garantiti dalle istituzioni governative attraverso norme e regolamentazioni che definiscano chiaramente quali sono i confini d'azione di tutti gli attori presenti all'interno dell'ecosistema digitale.



All'interno di questo elaborato ci si pone l'obiettivo di individuare quali siano le tendenze economiche relative alla condivisione e al mercato dei dati digitali in funzione dell'introduzione di alcune riforme normative (es. GDPR) applicate dalle istituzioni governative internazionali.

La struttura della tesi è così composta:

1. Literature review: si descriverà l'analisi effettuata sugli articoli e pubblicazioni presenti in letteratura relativi all'argomento in ottica di individuare le macroaree di studio della ricerca scientifica
2. Analisi delle ricerche e pubblicazioni effettuate in introduzione al modello sviluppato
3. Introduzione al dataset di riferimento e statistiche descrittive sulla base della domanda di ricerca
4. Sviluppo del modello analitico di regressione e riflessioni finali sulla base dei risultati ottenuti

2. Literature Review

2.1 Approccio metodologico

La literature review è stata effettuata nell'ottica di comprendere quali fossero i risultati della ricerca scientifica durante gli anni passati e identificare quali possano essere i trend e i gap di interesse all'interno del contesto del mercato dei dati e della digital privacy.

Si è iniziato con il definire quattro campi di applicazione della nostra literature review andando ad analizzare il contesto in cui si collocare l'elaborato:

1. Data market: si è preso in considerazione il mercato dei dati digitali come primo cluster di analisi
2. GDPR (*General Data Protection Regulation*): si è voluta analizzare la letteratura presente legata alla normativa pubblicata nel 2016 ma entrata in vigore nel 2018 relativa alla protezione dei dati digitali
3. Governance and regulation: si è analizzata la letteratura relativa ai campi applicativi della regolamentazione attinente al mercato dei dati digitali
4. Privacy: infine si è analizzata tutta la letteratura relativa alla privacy dei dati digitali all'interno dei mercati online

Qui di seguito in Fig. 1 si riporta uno schema sintetico di quello che è stato l'approccio metodologico utilizzato per le quattro categorie all'interno della literature review.



Figura 1: approccio metodologico della literature review.

Al fine di ottenere i risultati prefissati all'interno di questa fase si è previsto l'utilizzo di due strumenti dedicati alla ricerca scientifica e alla network analysis, in particolare:

- SCOPUS Database by Elsevier: è uno strumento che permette di effettuare ricerche sulla base di parole chiave ed individuare un cluster di documenti in funzione della ricerca effettuata. Inoltre, permette l'esportazione dei documenti individuati in modo da poterne analizzare il contenuto su un'altra piattaforma
- VOSviewer: è un software che permette di creare diverse tipologie di network map (es. co-citation map, co-occurrence map) con differenti visualizzazioni possibili su base sia cronologica che cromatica



2.2 Literature review: data market

Il primo campo di ricerca che si è voluto analizzare è quello relativo al mercato dei dati digitali (Data market): per fare ciò si è dovuto applicare il modello di ricerca descritto in Fig. 1 e adattarlo al campo applicativo.

In Fig. 2 si riporta il modello di ricerca metodologica applicato alla literature review per il mercato dei dati digitali.

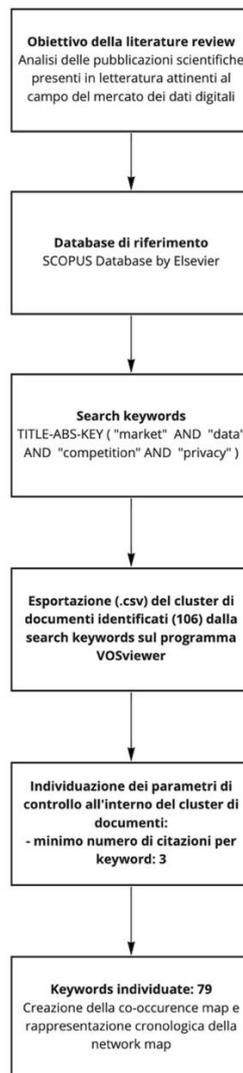


Figura 2: approccio metodologico alla literature review sul mercato dei dati digitali



- In basso a destra è riportata la legenda cromatica in funzione dell'evoluzione temporale dei campi di ricerca
- Ciascuna parola chiave è raffigurata in correlazione ad un singolo bullet: la grandezza è direttamente proporzionale al numero di citazioni riscontrate per quella mappa
- Gli archi rappresentano le correlazioni esistenti tra le varie keywords in funzione dei documenti del cluster

La creazione della mappa è stata funzionale all'analisi del cluster di documenti individuati dalla ricerca in SCOPUS, i quali coprono un orizzonte temporale dal 1984 al 2022 ma come si vede dalla mappa le keywords che rispettano la threshold di 3 citazioni sono comprese in documenti tra il 2014 e il 2020.

Dall'analisi dei documenti presenti all'interno del cluster esportato da SCOPUS si è visto come si è passati dalla necessità di determinare quale fosse il prezzo attribuibile al dato privato sul mercato digitale (Li, et al., 2013) capendone il vero valore in un'economia che stava nascendo e sviluppandosi enormemente con lo sviluppo delle big company in ambito digital. Parallelamente si è visto come il focus fosse su un'analisi economica di quello che era il mercato dei dati digitali che stava crescendo a velocità inaudite e che quindi andava compreso nelle sue peculiarità (Ding, et al., 2015).

Un altro campo che è stato notevolmente trattato è quello dell'AI (Artificial Intelligence) in quanto si è subito capito il potenziale del dato in ambito di sviluppo di nuove forme di intelligenza.

Dopo questa prima fase la ricerca scientifica è passata a studiare con notevole attenzione le applicazioni che il dato potesse avere nello sviluppo del mercato digitale online e come questo potesse aiutare le società che operano in questo settore ad individuare il consumatore target e ad influenzare le sue scelte veicolandolo verso il proprio prodotto (Jiao, et al., 2017). Il focus si è dunque spostato dall'analisi del valore del dato alla profittabilità del mercato connesso, andando a studiare quale potesse essere il vantaggio competitivo e il potere di mercato delle aziende operanti nel settore e come potesse essere mantenuto nel tempo.

La ricerca ha individuato sostanzialmente 4 categorie sull'Internet Value Chain" in cui andare ad analizzare il potere di mercato delle big company (Fast, et al., 2018):

1. Devices & App Stores
2. Search
3. E-commerce
4. Social Media



Al 2018 il potere di mercato big company riportava le suddivisioni raffigurate in Tab. 1, Tab. 2, Tab. 3, Tab. 4.

Tabella 1: Market Share, Devices & App Stores

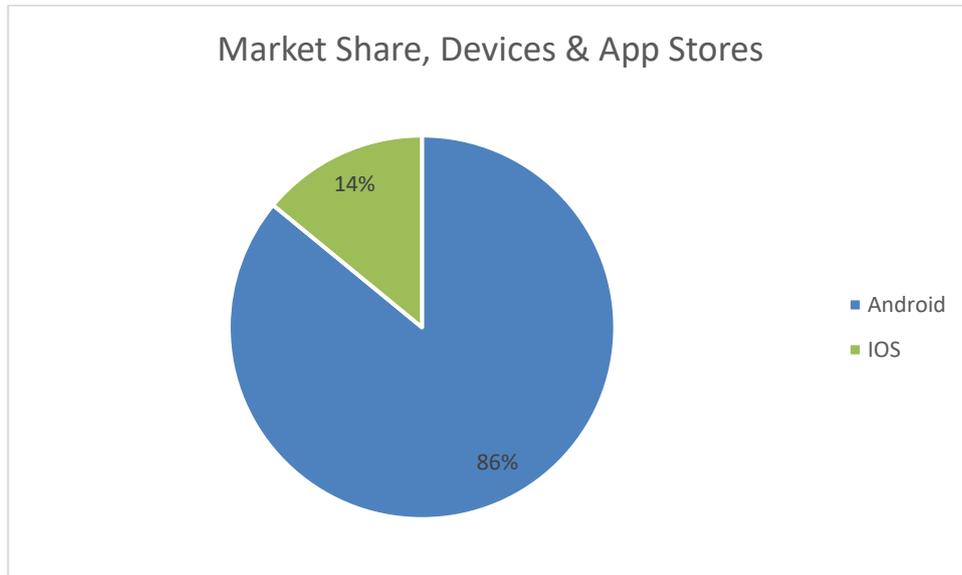


Tabella 2: Market Share, Search Industry

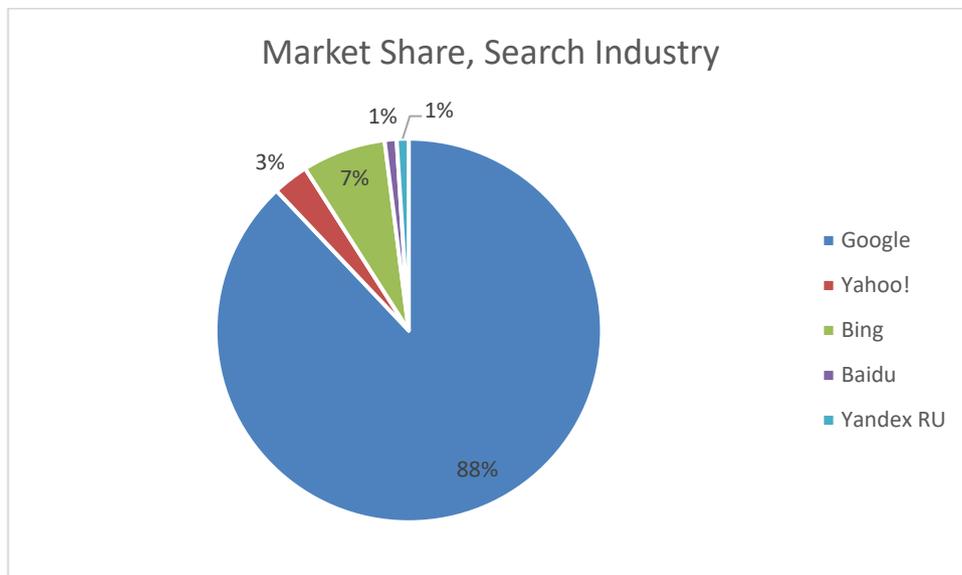
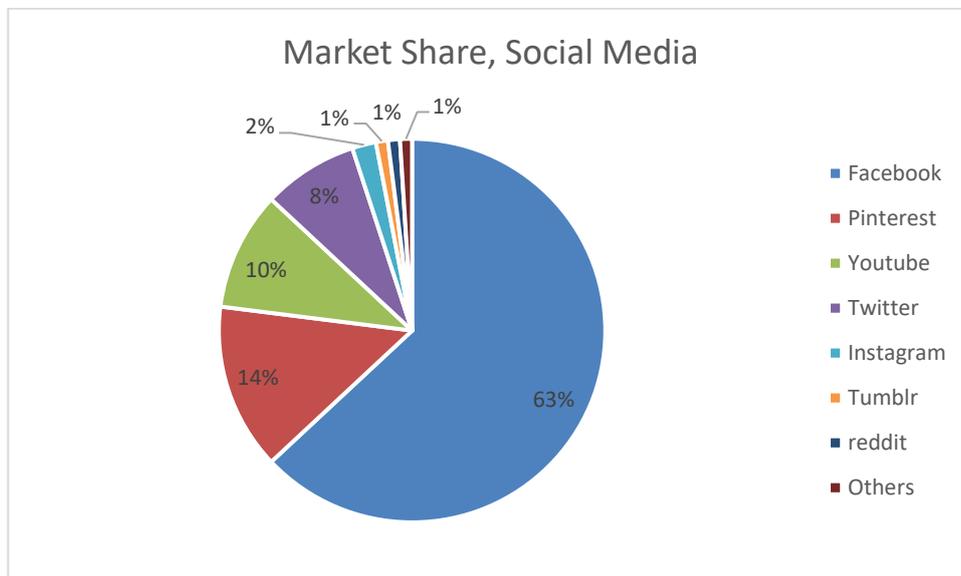




Tabella 3: Market Share, E-commerce



Tabella 4: Market Share, Social Media



Si è visto che l'utilizzo e la transazione dei dati personali si stava espandendo vertiginosamente con un processo di data collection che andava ad espandersi grazie



**Politecnico
di Torino**

all'avvento di nuove apps e tecnologie integrate come le Internet of Things correlate all'AI precedentemente citata (Acemoglu, et al., 2019).

Il risultato di molti dei documenti analizzati converge sempre ad un unico punto: i benefici derivanti dalla condivisione di dati e informazioni personali da parte dell'utente sono molteplici a partire dall'offerta di un prodotto/servizio maggiormente customizzato sui bisogni del cliente, una maggiore informazione relativa ai gusti e alle preferenze dei consumatori e infine la disponibilità di maggiori input per gli applicativi AI, andando a sfruttare le tecnologie emergenti della blockchain.

In conclusione, si va a definire in maniera chiara e netta quale sia realmente il ruolo del dato personale all'interno dei nostri mercati: il dato è inteso come input per lo sviluppo di applicativi e algoritmi al fine di tracciare le preferenze dei consumatori e veicolare il processo di decision making dei potenziali clienti.

È lecito pensare che tutti questi aspetti debbano comunque essere correlati ad un concetto di privacy estremamente chiaro e ben definito che vedremo essere stata studiata dalla ricerca scientifica approfonditamente nei capitoli successivi.



2.3 Literature review: GDPR

2.3.1 Introduzione al GDPR

Il secondo campo di ricerca che si è voluto analizzare all'interno della letteratura scientifica è stato quello relativo al GDPR (*General Data Protection Regulation*): tale regolamento è stato pubblicato in Gazzetta Ufficiale europea il 4 maggio 2016 ma è entrato in vigore il 24 maggio dello stesso anno per poi essere attuato a partire dal 25 maggio 2018.

Tale regolamento è composto da 99 articoli che si prepongono il raggiungimento dei seguenti obiettivi, nati da esplicite esigenze della stessa Commissione UE:

- Definitiva regolamentazione relativa alla protezione dei dati personali all'interno dei territori dell'UE
- Sviluppo del Mercato Unico Digitale europeo
- Risposta ai nuovi trend derivanti dalle nuove tecnologie digitali

Il fulcro principale del GDPR è lo switch dalla tutela dell'interessato alla responsabilità del titolare e dei relativi responsabili del trattamento stesso.

Si è tentato di sintetizzare nei punti seguenti quelli che sono i contenuti del GDPR e i punti cruciali relativi a questo regolamento:

- Con l'entrata in vigore del GDPR si sono introdotti importi ingenti per le sanzioni amministrative
- Si è introdotto il principio di accountability del titolare: ciò significa che *"il titolare e il responsabile del trattamento devono porre in essere tutte le misure di protezione effettiva al dato personale oggetto dei trattamenti"*.
Quindi diventa cruciale la valutazione dei rischi e la valutazione d'impatto che il titolare e il responsabile del trattamento devono effettuare.
Il principio di accountability prevede dunque l'adozione di comportamenti proattivi e l'esecuzione di azioni tali da dimostrare l'effettiva applicazione delle misure previste all'interno del GDPR
- Si è introdotto il concetto di "privacy by design" e "privacy by default": nel primo caso significa che il trattamento deve essere progettato nel rispetto dei requisiti del GDPR e nella tutela dei diritti degli interessati, nel secondo invece si stabilisce che si debbano rispettare i principi generali della protezione dei dati e che vadano quindi previste adeguate misure di sicurezza.
- Si sono previste regole maggiormente rigide per quanto riguarda la selezione e la nomina di un responsabile del trattamento
- Si è previsto l'ampliamento dei diritti che spettano al diretto interessato



- Si è inoltre previsto che il titolare del trattamento debba prontamente comunicare eventuali violazioni dei dati personali (data breach) direttamente al Garante

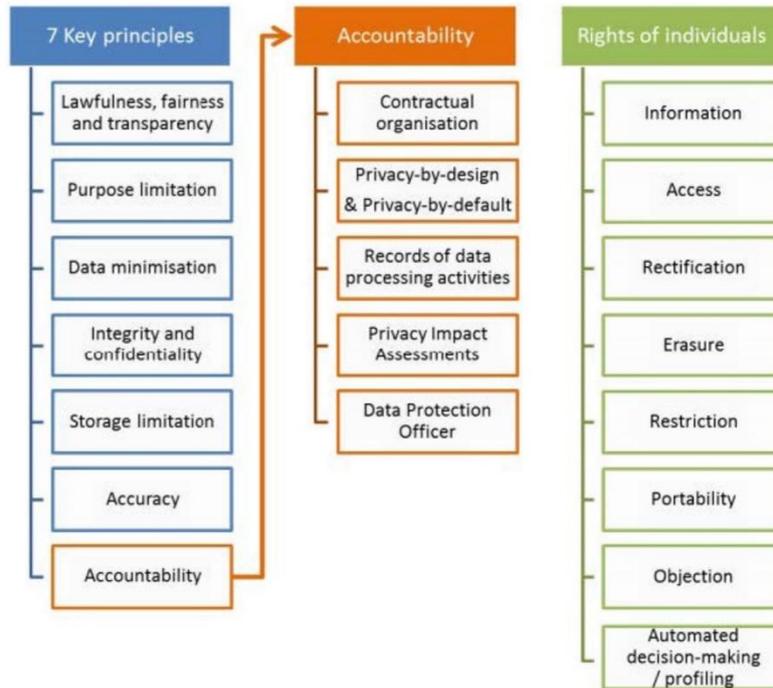


Figura 4: Mappa riassuntiva delle principali iniziative del GDPR (Carvalho, et al.)

2.3.2 Approccio metodologico per la literature review

Per analizzare la letteratura scientifica relativa alla produzione sul GDPR si è previsto un approccio leggermente differente rispetto a quello previsto all'interno del Cap. 2.1 per il "Data Market" in quanto risultava essere poco rilevante lo studio cronologico delle pubblicazioni a causa della recente entrata in vigore del GDPR.

Per tale motivo si è deciso di riportare all'interno della co-occurrence map (creata sempre con l'ausilio del software VOSviewer) una mappatura suddivisa per cluster cromatici e non più una mappatura impostata su scala cromatica cronologica.

Per ulteriori dettagli si rimanda ai paragrafi successivi.

In Fig. 4 si riporta il modello di ricerca metodologica applicato alla literature review per il GDPR.



Figura 5: approccio metodologico alla literature review sul GDPR

Come si può notare dunque dalla Fig. 4 è stata fatta una ricerca per parole chiave sulla base delle seguenti keywords: TITLE-ABS-KEY ("GDPR" AND "data").

Questo ha permesso di individuare un cluster di 2851 documenti², ridotto per difficoltà computazionali a 2000, che sono stati esportati in formato .csv e caricati sul software VOSviewer.

Una volta caricati il file su VOSviewer si è impostata una threshold relativa al numero minimo di citazioni per le keywords pari a 25 (molto più alta rispetto al caso analizzato al Cap. 2.1) individuando 102 parole chiave riportate all'interno della co-occurrence map in Fig. 5.

² si riporta il [link](#) al documento estratto da SCOPUS con i 2000 documenti analizzati per creare la co-occurrence map relativa al "GDPR"



5. In viola vengono riportate le tematiche relative alla macroarea di privacy in cui il GDPR si va a collocare

L'analisi della mappa è stata estremamente funzionale allo studio successivo che è stato fatto sui documenti estratti da SCOPUS che ha permesso di ottenere una visione completa ed esaustiva sulle conseguenze apportate, secondo la letteratura scientifica, dall'entrata in vigore del GDPR dal 25 maggio 2018 fino ad arrivare ai giorni nostri.

È chiaro che l'orizzonte temporale coperto dalla letteratura è molto più breve rispetto a quello analizzato al capitolo precedente data la recente introduzione del regolamento: per tale motivo i 2000 documenti estratti da SCOPUS coprono un orizzonte temporale che va dal 2019 al 2022, posteriore all'entrata in vigore del GDPR.

In particolare, si sono analizzati diversi documenti che hanno studiato dettagliatamente quelle che sono state le conseguenze relative all'introduzione del GDPR.

Si riportano dunque i risultati principali prodotti dagli studi effettuati dalla ricerca scientifica.

Si è visto come il GDPR abbia ridotto le richieste di dati personali da terzi nel breve periodo ma non si hanno ancora evidenze sul lungo periodo, per tale motivo non è possibile concludere che il GDPR abbia ridotto l'utilizzo di tecnologie di terzi da parte dei siti web che interagiscono con la privacy del consumatore. Inoltre, alcuni studi hanno evidenziato come il GDPR abbia influenzato la struttura e la composizione dei cookie che sono una specifica tipologia di richiesta: si è osservata una diminuzione dei cookie da terze parti ed una crescita parallela dei cookie raccolti direttamente.

Relativamente a questo punto si è notata una riduzione del 12,5% dei cookie mentre la tracciabilità del consumatore è cresciuta dell'8% (Aridor, et al., 2020).

Si è visto che il GDPR non ha solo influenzato i siti web direttamente connessi al regolamento ma ha influenzato enormemente il comportamento dei consumatori e il traffico di dati sui siti web che non sono direttamente soggetti alla normativa.

Infine, alcuni studi che hanno analizzato lo spostamento del potere di mercato sui portali online, hanno confermato quello che è già stato riportato in Tab. 2 ovvero che Google stia aumentando in maniera rilevante la sua quota di mercato anche a valle dell'introduzione del GDPR (Batikas, et al., 2020).

Uno dei settori più studiati dalla letteratura scientifica a valle dell'entrata in vigore del GDPR è stato quello medico: i dati medici dei cittadini, soprattutto dopo lo scoppio della pandemia e questo ha allargato notevolmente le prospettive del Garante in quanto sono stati condivisi e diffusi centinaia di miliardi di dati relativi alle condizioni di salute del singolo cittadino. Questo ha comportato diversi studi sulla correlazione esistente tra l'introduzione del GDPR e il settore medico, ad esempio, per quanto riguarda le politiche assicurative del singolo individuo.



**Politecnico
di Torino**

La tutela delle informazioni mediche personali del cittadino è uno dei topic cruciali che il Garante sta monitorando attentamente per il quale si prevedono ulteriori aggiornamenti futuri (Fatehi, et al., 2020).

Molto interessanti sono gli studi relativi ai trade-offs che esistono tra i requisiti richiesti dal GDPR e la tecnologia blockchain in quanto la tecnologia blockchain si basa sull'impermeabilità e sulla trasparenza che sono concetti fortemente discordanti con la protezione dei dati previsti dal regolamento.

Un altro punto di non contatto tra la tecnologia blockchain e i contenuti del GDPR sono le figure dei responsabili del trattamento dati contro la struttura a nodi decentralizzati della blockchain (responsabilità vs decentralizzazione).

È chiaro dagli studi effettuati che in futuro si prevederà di adattare la tecnologia blockchain a quelle che sono le esigenze dettate dalla privacy dei dati e si andrà a lavorare sui punti comuni esistenti tra queste due sfere apparentemente molto lontane (Tatar, et al., 2020).

In conclusione a questa seconda parte di literature review si può affermare che:

- l'introduzione della normativa relativa al GDPR abbia portato notevoli benefici ma ancora poco quantificabili su un orizzonte temporale di lungo periodo
- il GDPR comporti l'adattamento di molte tecnologie e settori alle sue linee guida
- Ci siano alcuni settori a cui bisognerà porre maggiore interesse (es. settore dell'healthcare)
- Le big company, nonostante l'introduzione di nuove restrizioni, stiano comunque continuando la loro crescita esponenziale (es. Google, Amazon)



2.3 Literature review: Governance and Regulation

Il terzo campo di ricerca al quale si è riposta attenzione è strettamente correlato a ciò che si è trattato all'interno del Cap. 2.2 ed è quello relativo alla Governance and Regulation: per fare ciò si è dovuto applicare il modello di ricerca descritto in Fig. 7, tornando ai principi del modello e delle mappe utilizzate nella literature review del Data Market, e adattarlo al campo applicativo.

In Fig. 7 si riporta il modello di ricerca metodologica applicato alla literature review per il campo di Governance and Regulation.



Figura 7: approccio metodologico alla literature review sulla Governance and Regulation



Come si può notare dunque dalla Fig. 7 è stata fatta una ricerca per parole chiave sulla base delle seguenti keywords:

TITLE-ABS-KEY ("antitrust" OR "governance" OR "monopolization" OR "regulation") AND TITLE-ABS-KEY ("privacy online" OR "market competition")).

È stata una costruzione molto più complessa quella della stringa di testo in input a SCOPUS in quanto bisognava correlare l'ambito Governance and Regulation con il contesto applicativo del mercato dei dati digitali e la privacy online.

Questo ha portato a produrre due differenti input che si sono correlati con l'operatore logico "AND".

Il procedimento ha permesso di individuare un cluster di 1008 documenti³ che sono stati esportati in formato .csv e caricati sul software VOSviewer.

Una volta caricato il file su VOSviewer si è impostata una threshold relativa al numero minimo di citazioni per le keywords pari a 7 individuando 144 parole chiave riportate all'interno della co-occorrenza map in Fig. 8.

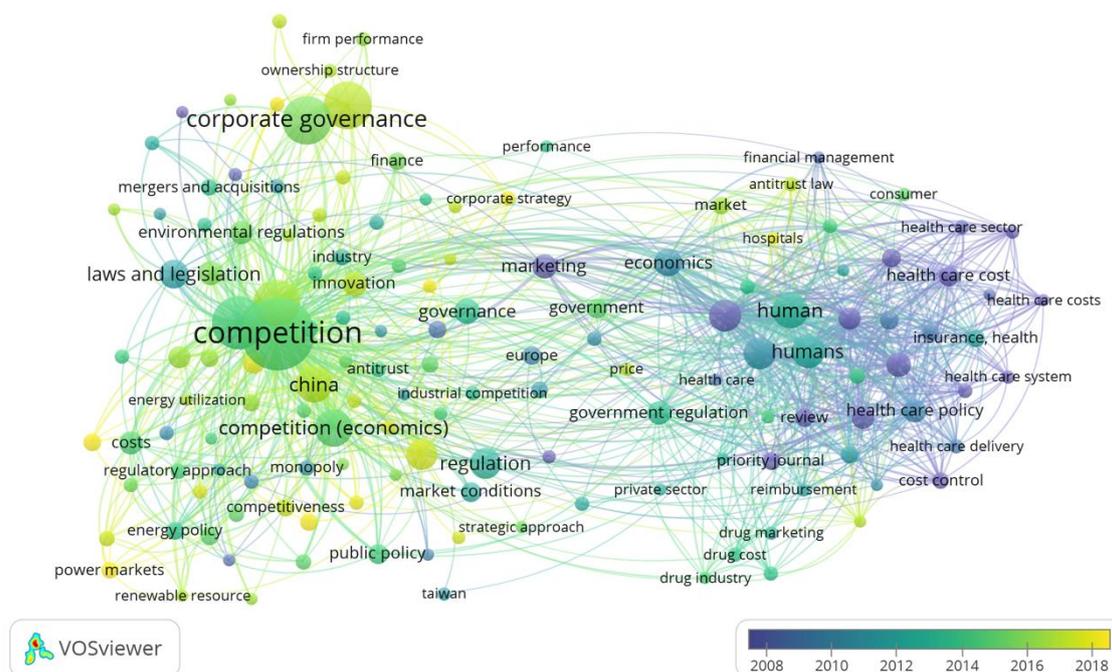


Figura 8: Co-occorrenza Map Governance and Regulation

³ si riporta il [link](#) al documento estratto da SCOPUS con i 1008 documenti analizzati per creare la co-occorrenza map relativa alla "Governance and Regulation"



Come si può vedere dalla co-occurrence map riportata in Fig. 8 la ricerca scientifica negli anni passati ha affrontato innumerevoli tematiche che vengono classificate a livello cronologico in due insiemi ben distinti, che si possono descrivere attraverso le seguenti osservazioni:

- Il primo cluster individuato dalla mappa è quello riportate in colore violaceo e rappresenta i temi trattati in anni meno recenti e come si può facilmente notare il tema principale è l'uomo inteso come consumatore e soggetto da tutelare all'interno della regolamentazione dei mercati e nel processo di monopolizzazione degli stessi
- Il secondo cluster, più recente, raffigurato con le parole chiave che vanno dal verde al giallo rappresenta il trend attuale della ricerca scientifica in campo di "Governance and Regulation". È facile comprendere come il tema centrale sia quello della competizione in quanto lo svilupparsi negli ultimi anni di colossi digitali come Amazon o Google, ha attirato notevolmente l'attenzione dell'antitrust in merito anche all'utilizzo lecito dei dati raccolti dai consumatori e alla tutela degli stessi.

Questa mappa ha permesso di ottenere una linea guida ben definita per lo studio dei documenti estratti da SCOPUS per comprendere quali siano e quali siano stati i pain point nel corso degli ultimi anni.

È doveroso comunque specificare che i documenti raccolti dalla ricerca in SCOPUS coprono un orizzonte temporale che inizia nel 1974 e termina ai giorni nostri, ma la threshold impostata sul numero di citazioni minime per le keywords ha ristretto l'intervallo temporale dal 2006 al 2022.

Partendo dallo studio delle prime pubblicazioni relative al ventesimo secolo si è visto come il focus centrale fosse il ruolo del consumatore e la sua tutela. In particolare, si era compreso immediatamente che si bisognasse dare la giusta attenzione all'aspetto medico e alle informazioni che le società potessero ottenere dalla condivisione dei dati personali (come riportato già nel Cap. 2.3).

Questo trend è stato poi traslato anche ai giorni nostri riportando diverse iniziative intraprese dalle big company nel mercato digital che hanno provato ad effettuare operazioni di fusioni e/o acquisizioni di altre società che operano in mercato in cui è molto frequente la condivisione dei dati medici da parte dell'individuo.

Non per ultima l'indagine messa in atto dall'Unione Europea nei confronti dell'operazione di acquisizione da parte di Google di Fitbit (Bourreau, et al., 2020).



Le preoccupazioni maggiori che sono sorte da questa indagine sono appunto legate al connubio che si verrebbe a creare tra la capacità di raccolta dati di Fitbit⁴ e il potere di mercato di Google: questo rischierebbe di comportare una notevole estensione del potere monopolistico del gigante statunitense e il conseguente aumento dello sfruttamento nei confronti del consumatore.

Questo proviene dal fatto che Google potrebbe combinare la mole enorme di dati che ha già a sua disposizione con i dati sanitari provenienti dal tracciamento dei devices di Fitbit; questo creerebbe opportunità per Google di entrare all'interno del mercato dei servizi assicurativi o di altre aree simili riducendo enormemente le asimmetrie informative con il consumatore, a suo favore.

Diventa quindi di vitale importanza il ruolo della Commissione Europea e di altre istituzioni nell'attuazione di azioni preventive per evitare l'accadere di fusioni e acquisizioni che possano accrescere ancora di più il potere di mercato dei colossi digitali e allo stesso tempo ridurre la libertà del consumatore.

Sempre legato all'aspetto sanitario è di notevole importanza studiare il ruolo che ha l'Intelligenza Artificiale (già citata precedentemente come fattore cruciale nel mercato dei dati digitali) rispetto al suo utilizzo nell'ambito di tecniche analitiche dei dati sanitari dei pazienti.

Attualmente non è presente un meccanismo di governance chiaro e definito per lo sviluppo di tali tecnologie in ambito sanitario; questo crea notevoli barriere all'ingresso nel mercato sia per l'utilizzo che per lo sviluppo dell'AI, ostacolando la concorrenza sul mercato.

Risulta quindi necessario adottare un meccanismo di regolamentazioni che tuteli l'applicazione dell'AI in ambito sanitario ma che allo stesso tempo tuteli la privacy del paziente e dei soggetti coinvolti (Morley, et al., 2022).

Risulta quindi essere ben definito il ruolo delle istituzioni internazionali come la Commissione Europea, la Federal Trade Commission o l'antitrust: sono gli attori che nel corso degli anni si sono occupati della tutela dei mercati e del consumatore.

È evidente come il loro lavoro sia andato ad intensificarsi negli ultimi anni dopo la crescita esponenziale delle piattaforme tecnologiche che hanno "oligopolizzato" il mercato e reso il ruolo del consumatore sempre più marginale.

Attualmente sono in atto diverse indagini da parte delle autorità sopra citate sulla liceità delle azioni intraprese dai colossi dei mercati digitali come Google, Amazon, Facebook ed Apple (Gibbons, et al., 2019), (Song, 2022).

⁴ Fitbit è una società statunitense che opera nel mercato dei "wearable device" che permettono il tracciamento dei parametri vitali del soggetto che li indossa



Dall'altro lato della medaglia però si devono analizzare anche gli aspetti benefici che la rivoluzione dei mercati digitali ha portato, in particolare l'avvento della digital economy (Subcommittee on Antitrust US, 2020).

Dal 2020, con l'avvento del COVID-19 si è compreso come l'accesso ad internet e quindi la conseguente condivisione di dati sia condizione necessaria e indispensabile e come le piattaforme precedentemente citate svolgano il ruolo di catalizzatore di questi meccanismi.

Ad esempio, Google è l'architettura cardine dello scambio di comunicazioni, beni e servizi nel mondo occidentale, senza il quale oggi l'economia mondiale non andrebbe avanti.

Negli ultimi anni però l'economia digitale ha manifestato una forte inclinazione alla concentrazione e alla monopolizzazione (si veda l'esempio dei social network): le prospettive nei prossimi dieci anni parlano di un 30% della produzione lorda mondiale a capo della big company della digital economy.

Questo processo di crescita così forte e repentino è dovuto principalmente al grande movimento di M&A che questo settore ha subito, creando dei colossi immensi che corrispondono in termini di produzione ad interi Paesi. Non sempre però le acquisizioni sono state volte all'inglobamento della società acquisita ma hanno previsto molte volte la dismissione della stessa ("killer acquisition").

In conclusione, i risultati di questa analisi possono essere riassunti nei seguenti punti:

- C'è necessità di regolare in maniera quanto più chiara e definita il mercato digitale
- Le autorità devono incrementare il loro raggio di azione e aumentare il numero di interventi all'interno di questo settore
- è di fondamentale importanza riconoscere e limitare il potere delle big company all'interno del mercato digitale, ma allo stesso tempo enfatizzare le esternalità positive che stanno avendo sull'economia globale



2.4 Literature review: Privacy

Il quarto ed ultimo campo di ricerca di cui si è voluta analizzare la produzione di articoli e pubblicazioni scientifiche è quello relativo alla Privacy: per fare ciò si è dovuto applicare il modello di ricerca descritto in Fig. 9, utilizzando i principi del modello e delle mappe utilizzate nella literature review del “Data Market” e della “Governance and Regulation”, e adattarlo al campo applicativo.

In Fig. 9 si riporta il modello di ricerca metodologica applicato alla literature review per il campo relativo alla Privacy.

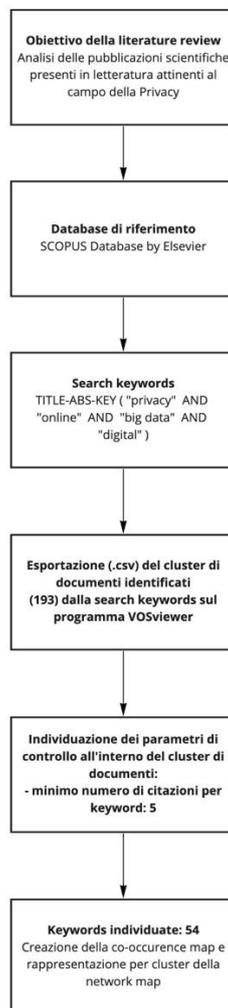


Figura 9: approccio metodologico alla literature review sulla Privacy



È opportuno sottolineare che questa literature review attinente alla macroarea della “Privacy” è stata condotta col fine di ottenere un risultato che riassume le evidenze estratte dalle tre precedenti literature review e che allo stesso tempo sottolineasse l’importanza di determinati aspetti che caratterizzano i trend attuali della privacy nel mercato dei dati digitali.

Anche se il campione temporale risulta essere ristretto ad un orizzonte di tre anni, l’analisi della co-occurrence map ha permesso di trarre diverse conclusioni che vengono riportate nei punti successivi:

- gli argomenti trattati dalla ricerca scientifica sono allineati con quanto descritto nei capitoli precedenti
- è evidente l’importanza e il ruolo centrale che hanno le tecnologie relative ai Big Data come data mining, cloud computing e blockchain
- è cruciale il ruolo della sicurezza all’accesso ai dati
- è necessaria la tutela del consumatore e dei dati personali che condivide
- è rilevante come le piattaforme di social network (es. Facebook) siano la fonte principale di discussioni, studi e riflessioni attinenti a questo campo

Vedendo più nel dettaglio l’evoluzione delle tematiche di ricerca negli ultimi anni si è voluto analizzare i documenti presenti nel cluster estratto da SCOPUS e si è visto come la ricerca scientifica sia partita dallo studio approfondito dei big data e dell’evoluzione dei social network (Thai, et al., 2016), (Verma, 2017).

Si è voluto partire dal provare a comprendere quale fosse una giusta definizione da attribuire ai big data che vengono introdotti dall’UE come “grandi quantità di tipi diversi di dati prodotti da varie fonti, fra cui persone, macchine e sensori” (Nicita A., 2020).

Nel corso degli anni le definizioni attribuite sono state tante ma l’aspetto univoco che ha legato queste definizioni sono appunto determinate caratteristiche che contraddistinguerebbero il mondo dei big data:

- velocità
- varietà
- volume
- valore

La connessione con il mondo dei social network è presto fatta: le piattaforme social sono il luogo al mondo in cui vengono scambiate il maggior numero di informazioni, ad una velocità inaudita, con tipologie estremamente variegata e dal valore “inestimabile” per le società che le utilizzano.

I big data infatti servono appunto per migliorare gli algoritmi con cui le società tracciano i gusti dei propri consumatori al fine di creare la giusta campagna di advertisement, ad esempio, per quel prodotto che rispecchia maggiormente i bisogni del consumatore stesso.

Questa è appunto la stretta correlazione che esiste tra il mondo dei big data e quello dei social network in ottica “Privacy”.



Dopo aver trattato di queste due tematiche estremamente importanti, la ricerca scientifica è passata a focalizzarsi sul ruolo delle tecnologie di blockchain applicate al mondo privacy (Kalan, 2020), (Peng, et al., 2020).

Anche in questo caso (come spiegato all'interno del Cap 2.2) il tema è proprio nel trade-off esistente tra la decentralizzazione e la totale trasparenza che la tecnologia blockchain ha introdotto all'interno del nostro sistema economico-sociale e la privacy dell'utente che viene ad essere scalfita dall'introduzione dallo sviluppo di tale tecnologia.

In conclusione a tutto il processo di literature review è necessario sottolineare l'importanza che ha avuto questo studio preliminare inerente al tema centrale della tesi in quanto ha permesso, attraverso questa fase di assessment iniziale, di:

- comprendere le criticità del mercato dei dati digitali e le prospettive future
- evidenziare gli attori principali all'interno di questo ecosistema e sottolineare l'importanza di ciascuno
- conoscere i cambiamenti dopo l'introduzione della normativa del GDPR
- capire quale sia il livello raggiunto dalla regolamentazione in materia e le eventuali lacune da colmare in un futuro prossimo
- carpire il raggio d'azione delle autorità competenti in tema antitrust e le azioni rilevanti su cui pongono la loro attenzione
- consolidare la posizione dominante delle big company all'interno di questo mercato



3. Studio di modelli applicati al mercato digitale

L'obiettivo di questo secondo capito è analizzare come le tecnologie connesse al mondo del web e le leggi inerenti alla regolamentazione relativa emesse negli anni hanno influenzato l'economia e le imprese, con un particolare focus a livello europeo, andando ad analizzare le differenze emerse con la realtà statunitense.

Possiamo suddividere i cambiamenti avvenuti in due tipologie in base alla loro causa:

- Dovuti all'introduzione di innovazioni, di cui la più rilevante è la banda larga ultraveloce
- Dovuti a cambiamenti nelle legislazioni e nelle normative, con l'introduzione della direttiva e-privacy prima e con il GDPR poi

In virtù dell'obiettivo che si è posto all'interno di questo elaborato si andranno ad analizzare con particolare attenzione le mutazioni avvenute a causa dell'introduzione nel 2016 della regolamentazione europea inerente al GDPR.

Tale regolamentazione ha portato i siti web a dotarsi di strumenti diversi per aderirvi, soprattutto riguardanti il tracciamento dei dati degli utenti.

Per svolgere la nostra analisi è quindi importante effettuare una panoramica dei cambiamenti avvenuti durante l'ultimo decennio e, per questo, sono stati selezionati ed analizzati, in aggiunta a quanto riportato nel Cap. 1, due articoli prodotti dal gruppo di ricerca del Politecnico di Torino capitanato dal Prof. Carlo Cambini (relatore della tesi) nel quale opera il Prof. Lorien Sabatino (correlatore della tesi).

I due articoli analizzati (Sabatino, et al., 2021), (Sabatino, et al., 2022) riguardano aspetti legali in particolare il primo si concentra sull'introduzione della regolamentazione inerente all'ePrivacy e al suo impatto sul mercato dei dati digitali; il secondo invece tratta l'introduzione del GDPR e il relativo impatto esteso a tutti i settori industriali senza focalizzarsi su specifici segmenti dell'economia.

L'articolo più rilevante è sicuramente l'ultimo che consente di fornire una panoramica e una descrizione dettagliate del dataset che verrà utilizzato nei capitoli successivi, per i modelli descritti inerenti alla comparazione che verrà effettuata, nello studio per l'impatto che il GDPR ha avuto, tra i siti Multibrand (i siti che commerciano e vendono prodotti di diversi marchi es. Amazon) e quelli Monobrand (i siti che commerciano prodotti monomarca es. Nike).



3.1 Online Privacy and Market Structure: Theory and Evidence (Sabatino L., Sapi G., 2021)

3.1.1 Introduzione

La direttiva ePrivacy ha introdotto in Europa regole rigorose per la gestione del traffico dati delle aziende operanti Online, per garantire al meglio la riservatezza delle informazioni personali degli utenti. Entrata in vigore nel 2002, la direttiva ha subito modifiche nel 2006 e nel 2009, quest'ultima in particolare ha avuto lo scopo di regolamentare ulteriormente acquisizione, stoccaggio e utilizzo dei cookie all'interno dei siti di eCommerce. In particolare, viene stabilito che l'utente debba avere a disposizione una spiegazione dettagliata del loro utilizzo e debba esplicitamente dare il proprio consenso. Un altro effetto di questa direttiva è stato la limitazione di spam e e-mail marketing, introducendo possibilità di sanzioni e richieste di risarcimento da parte del consumatore nel caso di ricezione assidua senza aver esplicitamente acconsentito. Per analizzare al meglio questa direttiva è necessario distinguere ciò che va a colpire: soltanto le attività aventi siti di vendita online e soltanto a livello europeo. Quest'ultimo fattore è di importante rilevanza ai fini dell'analisi in quanto permette un confronto dettagliato tra siti europei e statunitensi al fine di capire il reale impatto commerciale della direttiva.

L'analisi svolta nel documento mira a individuare le differenze portate a livello economico, facendo un confronto non soltanto con dati provenienti dall'esterno dell'Unione Europea, ma anche tra dati precedenti l'entrata in vigore e dati successivi, al fine di individuare trend e cambiamenti significativi.

3.1.2 Dataset

I dati utilizzati per l'analisi provengono da S&P CapitalIQ database, in particolare sono state selezionate aziende attive nel settore del retail online in Stati Uniti, Canada o Unione Europea. I prodotti venduti dalle imprese studiate sono beni di consumo discreti come vestiti, elettrodomestici o mobili, tre categorie di prodotti con elementi in comune ma con differenze riguardanti il rapporto tra la vendita online e la vendita fisica.

All'interno del dataset sono presenti dati di vendita relativi al periodo di tempo tra il 2003 e il 2017 in modo da coprire un periodo di tempo sufficientemente ampio da



valutare anche gli effetti a lungo termine della direttiva e non soltanto l'impatto di breve periodo.

Per ogni azienda sono oggetto di studio i dati relativi a:

- Ricavi
- Asset totali
- Asset correnti
- Stato operativo
- Classificazione dell'ambito di operatività

Un problema riscontrato nella selezione dei dati riguardava la localizzazione geografica in quanto la direttiva non si applica a livello mondiale, ma soltanto a livello europeo. Dato che molte società riportano dati soltanto a livello globale risultava difficile classificarli come dati di analisi (UE) o dati di controllo (NA), la soluzione è stata limitare il dataset a soltanto le aziende che riportavano i dati suddivisi per segmento geografico. Il risultato di questa selezione individua un dataset finale di 148 imprese.

3.1.3 Risultati

Un primo risultato osservabile riguarda la privacy offerta dalle imprese, è infatti dimostrato come quelle di dimensioni maggiori siano in grado di offrire ai propri clienti un livello di sicurezza più elevato, un dato confermato anche da siti web non appartenenti al settore della vendita online, dove si nota come i siti più grandi offrano mediamente gradi di protezione migliori.

In controtendenza rispetto a quanto detto sopra è riscontrabile, però che le imprese di grandi dimensioni hanno subito un impatto economico negativo derivante dalla direttiva più elevato rispetto a aziende più piccole, questo fatto può essere dovuto alla diversa capacità di monetizzare i dati raccolti da parte delle due categorie. Infatti, imprese piccole o medio-piccole hanno ricavi derivanti da dati raccolti molto più contenuti rispetto a imprese simili ma di grandi dimensioni e per questo anche a fronte di una riduzione di clienti maggiore derivante dalla minor privacy offerta, le loro entrate non hanno subito modifiche sostanziali. Al contrario le imprese di dimensioni rilevanti riescono a convertire ogni dato in guadagni più elevati e quindi anche a fronte di riduzioni di clienti più contenute (non si hanno dati relativi alle effettive riduzioni di numerosità ma è ipotizzabile questa differenza in virtù dei diversi gradi di sicurezza offerti) vedono una diminuzione dei loro ricavi più rilevante. Inoltre, per le imprese più penalizzate non è stata soltanto la perdita di clienti a portare questo calo, ma anche il



fatto che essere erano in grado, grazie ai dati raccolti, di influenzare le quote di mercato, con offerte personalizzate in base al cliente e altre strategie di marketing, e anche questa capacità ha subito un forte rallentamento a seguito della direttiva.

Ultima analisi riguarda invece una panoramica a livello macro del settore della vendita online, in questo caso non risultano effetti significativi di cambiamento né confrontando periodi di tempo diversi (antecedenti e seguenti alla direttiva), né confrontando paesi diversi (Stati Uniti e Canada rispetto all'Europa). Una possibile spiegazione è individuabile nel fatto che le imprese più colpite sono quelle di dimensioni rilevanti e quindi in numero minore, rispetto alla maggioranza del campione formato da aziende medio-piccole, portando così l'effetto di riduzione dei ricavi a non essere statisticamente significativo. Quest'ultimo risultato è visibile nel grafico seguente, dove NA indica gli stati del Nord America, utilizzati come confronto, mentre EU indica gli stati europei, si vede come il trend europeo non abbia subito modifiche rilevanti.

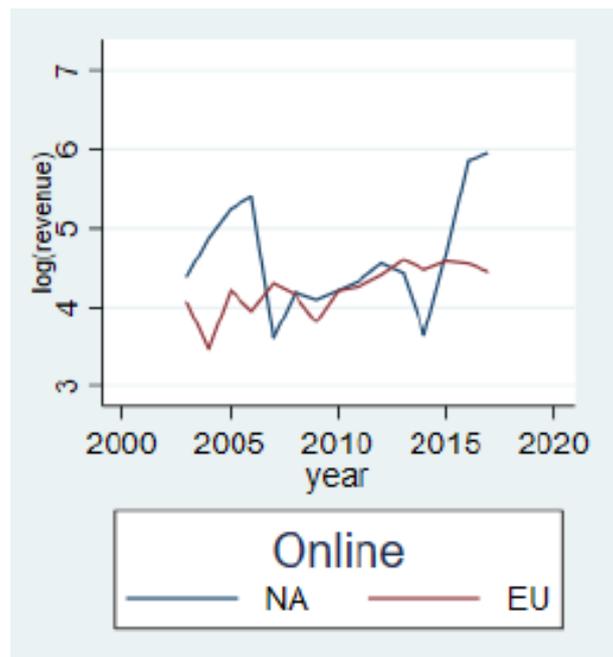


Figura 11: Trend di variazione revenues online market



3.2 The Impact of Privacy Regulation on Web Traffic Evidence From the GDPR (Sabatino L., Congiu R., Sapi G., 2022)

3.2.1 Introduzione

Conseguenza diretta della normativa ePrivacy già citata è stata l'emanazione del Regolamento Generale sulla Protezione dei Dati (GDPR), avvenuta nel 2018. La motivazione principale dietro questo documento è stata l'enorme crescita del mercato online (con tassi di crescita di gran lunga più elevati rispetto a quelli di ogni altro mercato), che ha portato i governi a preoccuparsi maggiormente di tutti i problemi legati alla privacy dei cittadini e a come i siti web utilizzano i dati raccolti. Il regolamento, valido per tutti i paesi appartenenti all'Unione Europea, stabilisce come le imprese che utilizzano questi dati debbano esplicitare all'utente tutte le modalità e debbano ottenere il consenso dello stesso prima di poterlo fare. In caso di utilizzo errato o non autorizzato vengono stabilite multe che possono arrivare fino al 4% del fatturato globale dell'impresa. Sebbene questo regolamento si applichi a tutte le tipologie di attività, coloro maggiormente colpiti saranno i siti web in quanto la loro principale monetizzazione deriva dalla capacità di raccogliere, aggregare e analizzare i dati dei propri utenti al fine di offrire loro pubblicità personalizzate, basti pensare al modello Google, il quale fonda la sua potenza sulla raccolta di dati relativi alle ricerche di chi lo utilizza, che vengono rielaborati al fine di proporre avvisi pubblicitari più vicini agli interessi di chi li riceve.

Sono ampiamente noti e studiati gli effetti che il regolamento ha avuto a livello di aumento del grado di protezione della privacy dei cittadini UE, ma materia di grande interesse sono anche gli effetti economici; infatti, il GDPR può influire su quattro aspetti distinti:

- Riduce le tecnologie di tracciamento utilizzabili (cookies, tracker, ...) dato che prima queste erano utilizzate senza che l'utente ne fosse messo a conoscenza, ed alcune di esse non sono più consentite perché troppo invasive e a basso grado di protezione;
- Riduce la pubblicità, in virtù sia dei costi maggiori per le aziende a creare annunci personalizzati sia dello stop ad alcune tipologie come quelle tramite e-mail;



- Aumenta il grado di rischio percepito dai siti web, in quanto stabilisce che questi condividano la responsabilità di eventuali infrazioni con i proprietari degli annunci;
- Porta all'attenzione di un pubblico sempre maggiore tutti gli argomenti legati alla privacy e ai nuovi limiti imposti, creando consapevolezza in un numero maggiore di utenti web che andranno a leggere più nel dettaglio le condizioni di trattamento dei dati, e in molti casi a rifiutarle.

Dal punto di vista giuridico dopo l'entrata in vigore del GDPR c'è stato un periodo di assestamento in cui non veniva ancora applicato alla lettera, che è terminato il 21 gennaio 2019 con la sanzione di 50 milioni di euro inflitta a Google per la mancanza di trasparenza e la mancanza di valido consenso per gli annunci. Questa sanzione rappresenta un passo molto importante, perché viene inflitta a un'azienda di grande rilevanza mediatica e internazionale, ed ha sicuramente influenzato molte altre aziende portandole verso un adeguamento delle loro misure di sicurezza.

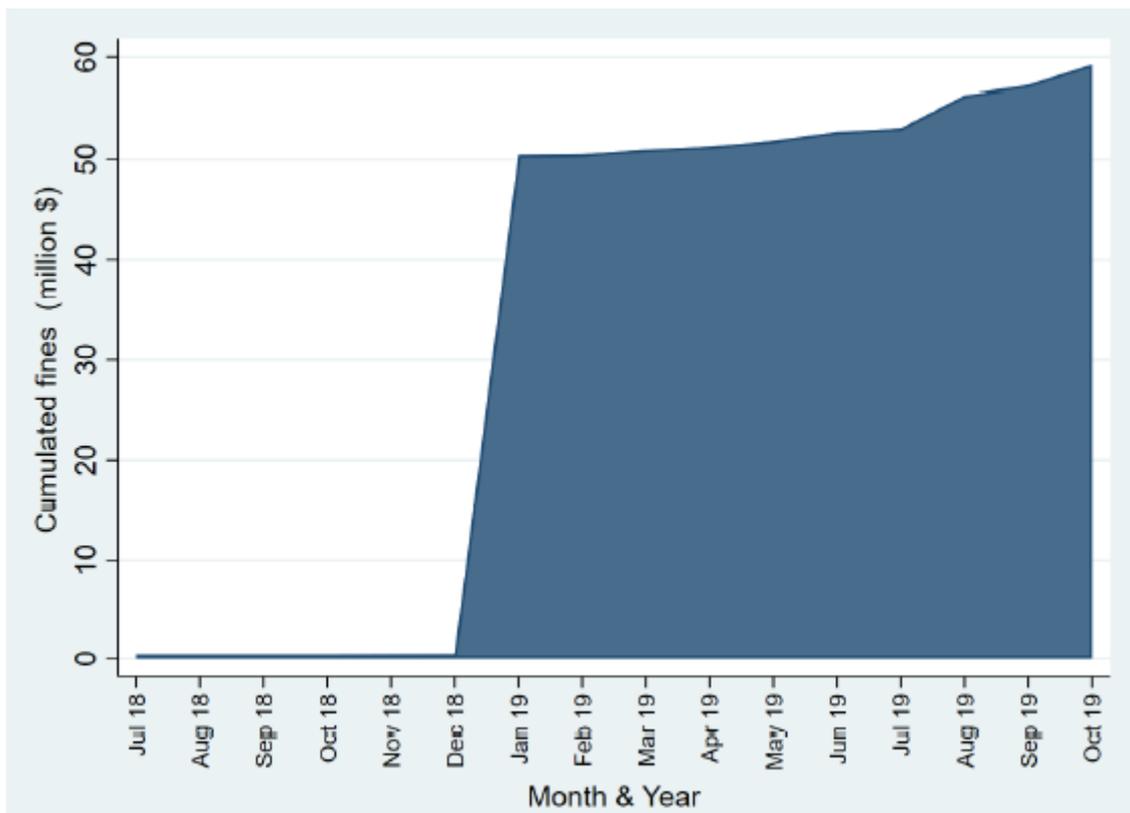


Figura 12: Cumulata delle sanzioni dovute al GDPR



Inoltre, come possiamo vedere nel grafico, alla prima sanzione di grande impatto (il salto è dovuto proprio ad esse) sono seguite altre sanzioni con impatto minore, questo indica come i legislatori non siano più disposti a “chiudere un occhio” sul rispetto delle normative.

L’obiettivo dell’articolo in esame è individuare e modellizzare come il GDPR abbia influito sul traffico online, confrontando i siti sia nel tempo (prima e dopo l’entrata in vigore) sia per la tipologia e le dimensioni dei siti web.

3.2.2 Dataset

La fonte principale per i dati utilizzati, presenti all’interno del dataset, è un online data provider che per i siti web presenti nel suo database registra:

- **Visite:** il numero di visite totali a un certo dominio (anche se un utente visita più pagine dello stesso sito conta come una sola visita) nel periodo di tempo indicato;
- **Durata media della visita:** cioè la media delle durate delle visite di un dominio in secondi;
- **Pagine medie per visita:** il numero medio di pagine del dominio che vengono visualizzate per ogni visita;
- **Frequenza di rimbalzo (Bounce Rate):** il numero di utenti che esce dal sito senza interagire con esso (ad esempio visualizzando pagine o contenuti) dopo esservi entrato;

I dati all’interno del dataset vengono riportati con:

- Il riferimento cronologico del dato estratto;
- Il Paese di origine della visita effettuata sul sito di riferimento;
- L’industria relativa al dominio di riferimento (settore industriale di appartenenza)

Grazie ai dati finora descritti è possibile svolgere un’analisi temporale, cioè valutare come si è modificato il comportamento degli utenti nei confronti dello stesso sito prima e dopo l’entrata in vigore del regolamento.

Inoltre, grazie alla distinzione per paese di provenienza è possibile valutare come utenti soggetti al GDPR e non interagiscano con lo stesso sito web. In questo modo si svolge sia un’analisi di trend temporale sia un confronto puntuale, minimizzando il rischio di ottenere risultati devianti o in qualche modo influenzati da eventi esterni.

Per lo studio vengono selezionati i siti web attivi negli Stati Uniti e nei cinque paesi principali europei: Francia, Italia, Germania, Paesi Bassi e Regno Unito, per un totale di 4957 siti attivi prima e dopo l’entrata in vigore del regolamento.

I dati coprono un periodo temporale di 24 mesi totali, 7 precedenti e 17 successivi all’entrata in vigore, dal 1° novembre 2017 al 31 ottobre 2019.

Per far fronte all’alta volatilità giornaliera i dati sono aggregati su base settimanale, in questo modo si ottiene un totale di circa **quattro milioni di osservazioni**.



Un'ulteriore suddivisione viene effettuata per canale attraverso cui il dominio viene raggiunto:

- **Direct:** quando l'utente digita direttamente l'indirizzo IP nella barra degli URL del browser, oppure quando clicca su link esterni al browser che lo indirizzano al sito (non ricevuti via mail)
- **Display ads:** visite successive a click su banner pubblicitari o advertising online
- **E-mail:** accessi a seguito di link ricevuti tramite mail
- Organic Search, quando si visita un sito dopo averlo trovato tramite un motore di ricerca come Google o Bing
- **Paid search:** simile alla precedente ma il click avviene su annunci a pagamento proposti dal motore di ricerca
- **Social:** accessi grazie a link presenti e pubblicizzati sui social network come ad esempio Facebook oppure Instagram
- **Referrals:** visite successive a click su collegamenti presenti in blog, forum e tutto ciò che non rientra nelle categorie precedenti.

Questa suddivisione permette di valutare quale canale sia stato influenzato maggiormente dal GDPR, se l'effetto sia stato omogeneo oppure no.

Possiamo distinguere ancora i canali in due categorie, il traffico a pagamento e quello gratuito, del secondo fanno parte soltanto il canale diretto e la ricerca organica, in quanto tutti gli altri generano costi per l'impresa. Ad esempio, pagando per pubblicità su Google o Facebook o acquisendo elenchi di indirizzi e-mail a cui inviare da terze parti specializzate.

Il grafico seguente (Fig. 13) mostra la suddivisione percentuale delle visite in base al canale da cui provengono. È evidente come la maggior parte provenga dal canale non a pagamento, circa l'87%, e come la ricerca a pagamento o gli ads rappresentino una parte quasi trascurabile del traffico totale, pur generando introiti rilevanti per le società che mettono a disposizione di clienti le pubblicità online.

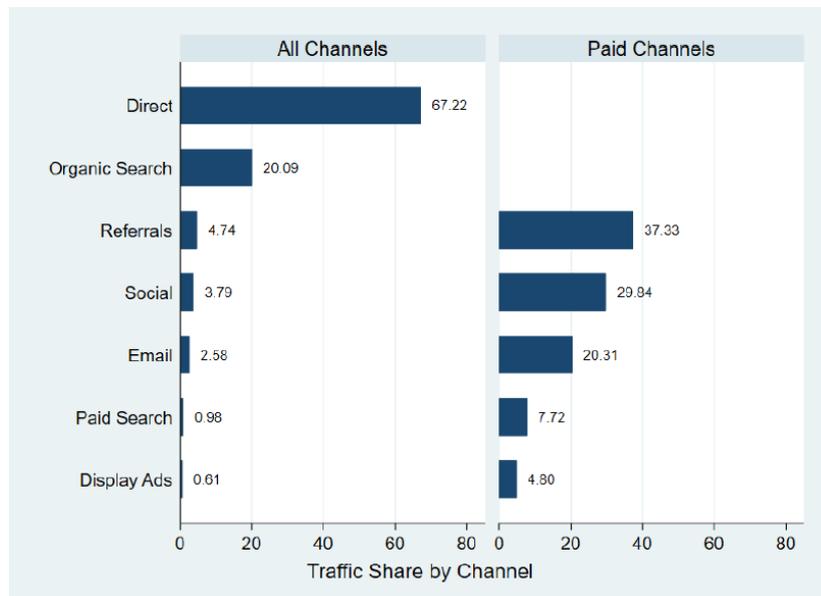


Figura 13: Distribuzione del traffico online per canale

Un possibile problema nei dati poteva riguardare il fatto che a fronte delle norme presenti nel regolamento, il data provider non fosse in grado di fornire valori con la stessa precisione precedenti in virtù del fatto che i siti non potessero inviare i loro dati senza esplicito consenso degli utenti. Questa ipotesi è, però, stata scartata dopo aver effettuato tre test:

- È stato esplicitamente chiesto al data provider se potesse esserci una potenziale distorsione nei dati, la risposta scritta è risultata essere negativa;
- È stato valutato come i dati utilizzati dal dominio non provengono soltanto dai siti in esame ma anche da altre fonti come fornitori di rete, fonti di dati pubbliche e altre tipologie, molte delle quali non identificano individualmente i dati ma solo in aggregato e quindi non sono coinvolte dal GDPR;
- In ultimo, è stato effettuato un confronto tra i dati presenti nel database e quelli di un erogatore di servizi simili operante in Germania per una categoria di dati non interessati dalle nuove normative e non sono state rilevate deviazioni significative tra le due fonti.

In conclusione, possiamo assumere i dati come idonei all'analisi.

3.2.3 Risultati

Il modello sviluppato evidenzia un effetto negativo statisticamente significativo in conseguenza all'entrata in vigore del GDPR. Infatti, il traffico dati si è ridotto in media del 15%, questa riduzione non si è però verificata subito dopo l'emissione ma è iniziata durante il periodo in cui come detto non venivano ancora applicate le sanzioni e si è



assestata sul valore riportato circa un anno dopo (Fig.14). Questa statistica temporale è valida per ogni tipo di canale di acquisizione.

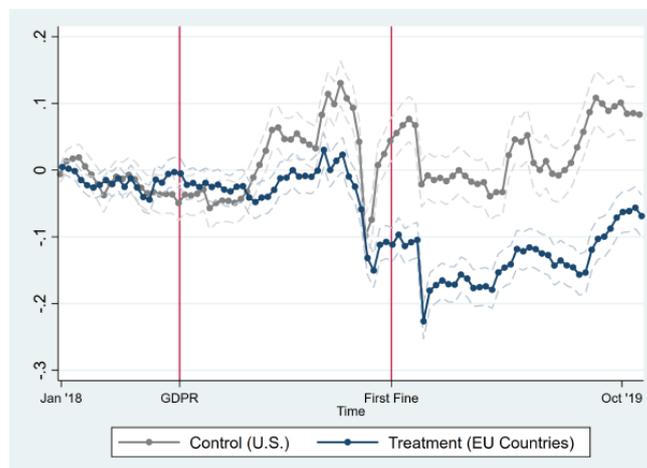


Figura 14: Variazione del traffico web

Nel breve periodo i canali più penalizzati sono stati: E-mail con una riduzione dell'8% e il canale diretto (4%).

L'analisi di lungo periodo invece porta una riduzione del 35% per le visite portate dalle e-mail e del 29% per le pubblicità su display, identificando questi come i canali colpiti più duramente. In realtà per capire al meglio questo dato bisogna puntualizzare che soltanto il 3% del traffico di un sito web deriva da queste fonti; quindi, anche una riduzione del numero di visite più contenuta porta una variazione percentuale rilevante. Per quanto riguarda gli altri canali tutti presentano riduzioni statisticamente rilevanti fatta eccezione per la ricerca a pagamento, dove non risulta significativo né a breve né a lungo termine. Anche il canale diretto subisce una riduzione, sebbene più contenuta e pari a circa il 7%. Quest'ultimo dato, però, potrebbe avere una base diversa, infatti non rientra in canali pubblicitari, ma la diminuzione potrebbe essere dovuta alla maggiore consapevolezza generata negli utenti, grazie ai banner che i siti web devono mostrare agli utenti per spiegare l'utilizzo dei loro dati effettuato.

La seconda analisi riguarda l'effetto generato sui siti classificati in base alla loro dimensione, è giusto premettere che la maggioranza (59%) di essi sono di piccole dimensioni e generano soltanto il 6% del traffico totale, i siti medi (30%) generano il 13% del traffico totale, i grandi siti (10%) generano il 23% e i colossi del web (1%) generano la gran parte, cioè il 57% delle visite. Per queste riduzioni percentuali simili possono portare a valori assoluti molto diversi tra loro.

Nel breve la riduzione risulta significativa e con un valore pari al 14% soltanto per i siti di piccole dimensioni. Nel lungo al contrario si verifica una relazione a forma di U rovesciata, cioè i siti piccoli e grandi subiscono variazioni significative, mentre i siti di



medie dimensioni risultano quasi inalterati. Questa relazione presenta comunque delle differenze rilevanti, infatti, per i siti piccoli si assesta su un valore di circa il 41%, per i grandi sul 7% e per i colossi 16%. La statistica trovata sembra confermare un'opinione politica molto diffusa durante la discussione precedente l'emissione del GDPR che sosteneva che questo avrebbe avuto un impatto devastante sulle piccole imprese.

L'ultima analisi riguarda tutte le altre statistiche riportate in precedenza:

- Le durate medie delle visite subiscono una riduzione del 3,5% nel lungo periodo;
- Il numero medio di pagine visitate cala del 1,5%;
- La frequenza di rimbalzo aumenta dello 0,6%, un valore basso ma non irrilevante.

Questi tre risultati dimostrano come non siano diminuite soltanto le visite, ma anche il coinvolgimento degli utenti, in particolare dopo l'emissione della prima grande multa citata in precedenza.

In conclusione, si può affermare come la maggiore consapevolezza e documentazione fornita agli utenti abbia portato un atteggiamento diverso nei confronti dei siti web e, in termini economici, renda la pubblicità sempre più nelle mani dei grandi fornitori di servizi, come Google, in quanto l'unico canale che non presenta riduzioni significativamente rilevanti è quello della ricerca a pagamento.

Non si possono, però, trarre altre conclusioni di tipo economico in quanto, benché la monetizzazione sia legata principalmente alle visite, gli utenti residui essendo maggiormente consapevoli potrebbero portare ad una monetizzazione maggiore in quanto più propensi a visualizzare banner pubblicitari.

Nel complesso si può affermare, comunque, che i vantaggi in termini di privacy portati ai consumatori superino gli effetti negativi creati nella struttura del web market, a fronte di politiche meno invadenti e più chiare intraprese dai siti.



**Politecnico
di Torino**

3.3 Conclusione

Dai due articoli descritti è evidente come negli ultimi anni ci sia stata una rapida evoluzione del mondo digitale che ha avuto effetti lungo tutta la catena produttiva, partendo dalla creazione dei prodotti, fino alla loro vendita e distribuzione finali, in particolar modo nella fase pubblicitaria. Sarebbe possibile individuare molti altri articoli riguardanti tematiche simili, tutti portanti evidenze complementari. In particolare, si può citare un articolo “The Faster the Better? Advanced Internet Access and Student Performance” che studia la connessione tra la banda larga ultraveloce e l’apprendimento degli studenti, che potrebbe avere analogie con l’apprendimento dei lavoratori delle imprese quando emergono nuove procedure legate al mondo online.

A livello economico si può individuare un cambiamento degli impatti: infatti, se per la normativa sull’ePrivacy l’impatto più rilevante è stato sulle imprese di dimensioni maggiori, l’introduzione del GDPR ha colpito in particolar modo le imprese più piccole. Questo risultato è in linea con le tendenze economiche che si stanno verificando a livello globale, dove risulta un lento ma costante spostamento dei clienti dalle piccole imprese ai colossi dei settori. Questi due valori, generale e online, possono avere una base comune, cioè la capacità di generare economie di scala e di apprendimento che portano le imprese più grandi a offrire prezzi più competitivi oppure servizi di migliore qualità.



4. Introduzione al dataset di riferimento e statistiche descrittive sulla base della domanda di ricerca

4.1 Domanda di ricerca

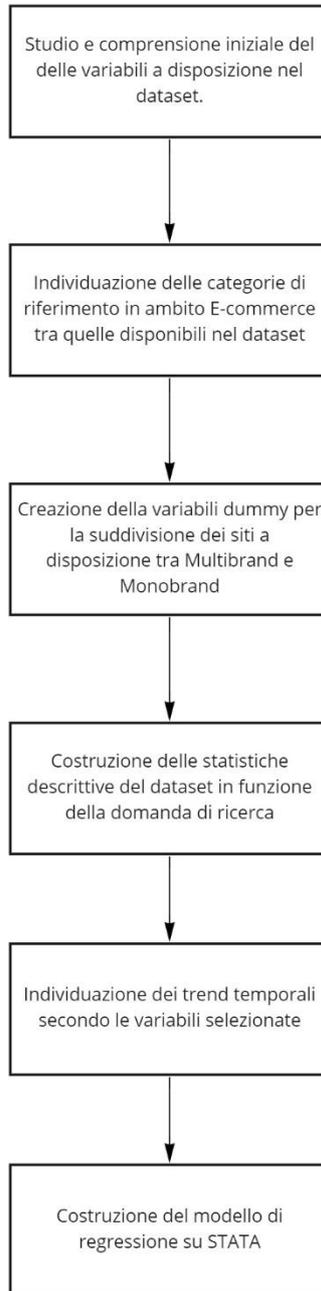
Dopo aver analizzato la produzione della letteratura scientifica inerente agli ambiti del mercato digitale, GDPR, Governance and Regulation e Privacy e gli articoli in ambito prodotti dal gruppo di ricerca dai Prof. Carlo Cambini e Lorien Sabatino, in modo da avere una prospettiva sugli studi già effettuati sul dataset a disposizione, si è deciso di focalizzare l'attenzione di questa tesi sulla valutazione dell'impatto che l'introduzione del GDPR ha avuto sui siti commerciali Multibrand e Monobrand.

Lo studio prevede l'individuazione nel dataset dei siti di riferimento attraverso l'assegnazione di una nuova variabile dummy che discretizzerà le due categorie all'interno dei vari settori industriali a disposizione e successivamente l'analisi dei trend temporali, nello spazio temporale a disposizione dei dati, attraverso l'impiego di Excel e di un modello di regressione su STATA.

Si è deciso di porre l'attenzione su questa tematica in modo da comprendere quali siano stati gli impatti del GDPR, dalla data di entrata in vigore, nel settore dell'E-commerce tra i colossi mondiali del **multimarket** come Amazon.com e Alibaba.com e colossi **monomarca** Nike.com o Adidas.com. Inoltre, si è cercato di sottolineare le differenze riscontrate anche da un punto di vista geografico, in quanto si aveva a disposizione il traffico extra-europeo, in particolare degli Stati Uniti, in cui la normativa inerente al GDPR non è prevista nella giurisdizione nazionale.



Lo studio è stato effettuato seguendo la procedura che viene riportata attraverso il workflow in Fig. 15.



mira

Figura 15: Procedura metodologica applicata per lo studio proposto



4.2 Introduzione al dataset e statistiche descrittive

Il dataset utilizzato per lo studio della domanda di ricerca introdotta all'interno del Cap 4.1 è analogo a quello utilizzato nell'analisi riportata all'interno del Cap 3.2.

I dati all'interno del dataset fanno riferimento ad un arco temporale che inizia il 30 Dicembre 2017 e termina il 25 Ottobre 2019; per comodità di analisi i dati sono stati riportati con frequenza settimanale andando ad aggregare i valori registrati sull'ultimo giorno disponibile della settimana di riferimento.

I dati vengono suddivisi in 15 colonne, che riportano:

1. **"Week"**: la settimana di riferimento in cui si è osservato il traffico
2. **"Domain"**: il nome del sito web
3. **"Country"**: il Paese di provenienza della visita sul sito web riportato nella colonna "Domain"
4. **"Full Category"**: il settore industriale specifico
5. **"Category"**: il segmento economico di riferimento
6. **"Traffic"**: il traffico medio della settimana
7. **"Avg_Visit_Duration"**: la durata media della visita
8. **"Avg_Pages_Visited"**: il n° medio di pagine visitate
9. **"Avg_Bounce_Rate"**: la frequenza di rimbalzo media dell'utente
10. **"OverallVisits"**: il numero totale di visite
11. **"OverallAvgVisitDuration"**: la durata media totale delle visite
12. **"OverallAvgPagesVisited"**: il numero medio totale di pagine visitate
13. **"OverallBounceRate"**: la frequenza di rimbalzo totale
14. **"OverallPageViews"**: numero totale di visualizzazioni per pagina
15. **"CountryTrafficShare"**: la percentuale del traffico del sito proveniente dallo Stato di riferimento, rispetto al totale del traffico del sito

Per ciascuna colonna sono presenti 620.000 dati, per un totale di 9,3M di voci complessive.

Per perseguire l'obiettivo posto all'interno della domanda di ricerca si è ristretto il dataset rispetto alle categorie e al settore industriale di riferimento e si sono definite delle variabili aggiuntive (Dummy) per discretizzare i dati secondo ulteriori criteri utili ai fini della tesi.



La prima attività effettuata è stata individuare le categorie di interesse tra quelle proposte all'interno del dataset, in particolare, i dati del traffico sono riportati per siti suddivisi nelle seguenti categorie:

1. Arts_and_Entertainment
2. Computers_Electronics_and_Technology
3. Games
4. Health
5. Reference_Materials
6. News_and_Media
7. Business_and_Consumer_Services
8. File_Sharing_and_Hosting
9. Science_and_Education
10. Hobbies_and_Leisure
11. Sports
12. Food_and_Drink
- 13. E-commerce_and_Shopping**
14. Adult
15. Finance
16. Home_and_Garden
- 17. Lifestyle**
18. Travel_and_Tourism
19. Heavy_Industry_and_Engineering
20. Law_and_Government
21. Vehicles
22. E-commerce_Related
23. Jobs_and_Career
24. Community_and_Society
25. Gambling
26. Universities_and_Colleges
27. Search_Engines
28. Pets_and_Animals
29. Social_Networks_and_Online_Communities
30. Email

Lo studio proposto all'interno di questa tesi si è concentrato appunto sul traffico dei siti provenienti dalle categorie di "E-commerce and Shopping" e "Lifestyle". In particolare, all'interno di queste macrocategorie si sono individuati i soli siti appartenenti al seguente elenco di settori industriali:

1. Lifestyle/Fashion_and_Apparel
2. E-commerce_and_Shopping/E-commerce_and_Shopping



3. E-commerce_and_Shopping/Marketplace

A valle di questa prima scrematura del dataset, che ha portato all'individuazione di 289 siti web, è stata necessaria la creazione di una variabile dummy che andasse a suddividere i siti web presenti all'interno di queste categorie tra quelli Multibrand e quelli Monobrand.

Questa attività è stata effettuata grazie all'utilizzo di informazioni estrapolate dagli stessi siti internet e dalla piattaforma data provider, che fornisce molte informazioni una volta dato in input il dominio del sito di cui si vogliono conoscere le informazioni.

In particolare, si sono individuati (si riportano la suddivisione numerica e percentuale in Tab.5 e Tab.6):

- 226 siti web per la categoria Multibrand
- 63 siti web per la categoria Monobrand

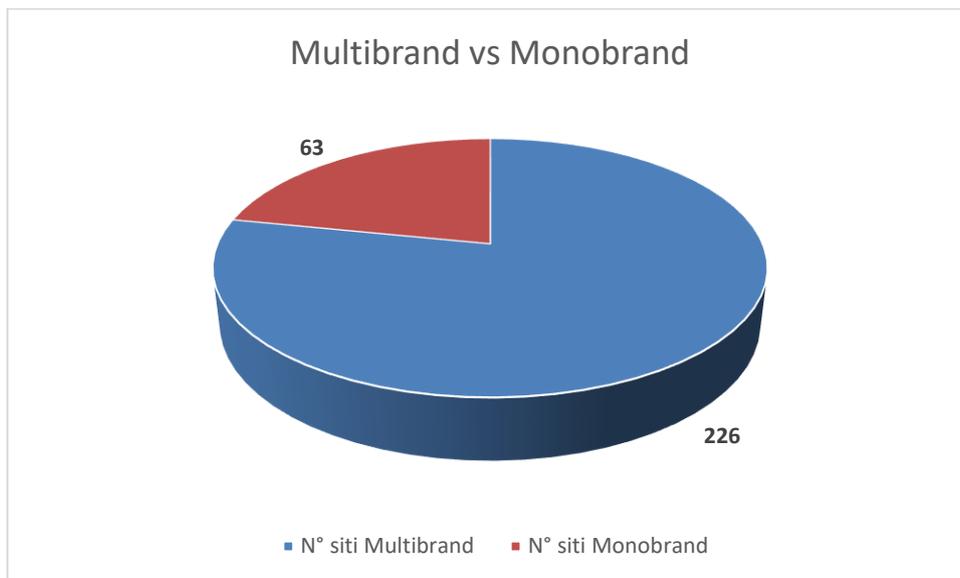


Tabella 5: Suddivisione siti Multibrand vs Monobrand

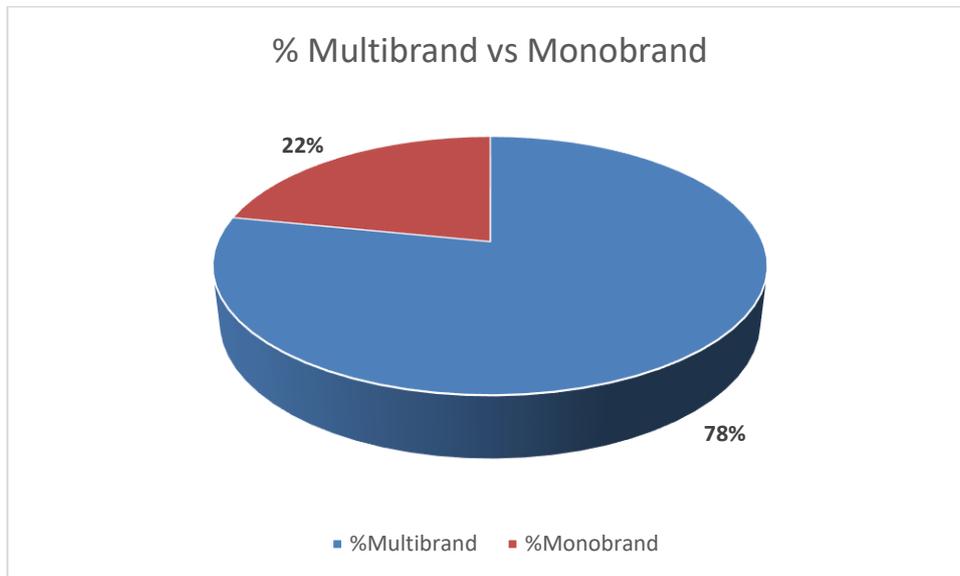


Tabella 6: Suddivisione percentuale siti Multibrand vs Monobrand

Come si può notare si ha a disposizione un campione più grande di siti Multibrand rispetto a quelli Monobrand, nonostante il numero di siti Monobrand sia comunque rilevante. In Tab. 7 e in Tab. 8, riportate in Appendice A, si evidenziano l'elenco dei siti Multibrand e di quelli Monobrand individuati attraverso la creazione della variabile dummy.

La creazione della variabile dummy è stata fondamentale per il proseguo dell'analisi in quanto ha permesso di individuare i due cluster di riferimento per lo studio del traffico.

Successivamente, una volta individuate le due macrocategorie, si sono effettuati degli studi per comprendere maggiormente quali fossero le caratteristiche di questi siti; in particolare si è condotto uno studio per l'identificazione dei siti di maggiore rilevanza in termini di visite e uno studio per la segmentazione geografica delle ricerche.

Partendo dalla prima segmentazione in funzione della dimensione dei siti, si è applicata la procedura riportata in Fig. 16.

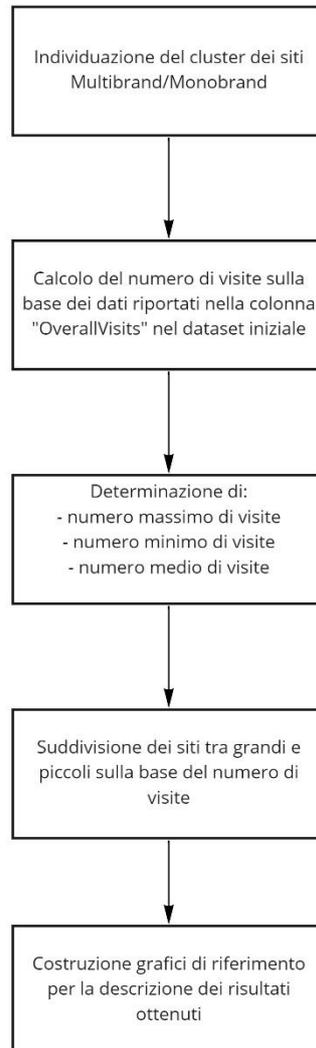


Figura 16: Procedura per l'individuazione della suddivisione dei siti sulla base della loro dimensione in termine di n° di visite totali

Come si può evincere dalla Fig. 16, il primo passo è stato quello di individuare, per i siti precedentemente suddivisi nelle due categorie (Multibrand e Monobrand), il n° di visite complessive effettuate sul periodo di riferimento.

Si è presa la colonna del dataset riportante il numero di visite complessive per ciascuna settimana e per ciascun sito e si è calcolata la somma per tutte le settimane di riferimento fino all'entrata in vigore del GDPR il 25/05/2018 per non avere influenze sulle dimensioni dei siti dall'entrata della normativa.



In questo modo si sono ottenuti i valori relativi al n° di visite complessive per ciascun sito, nel periodo di riferimento. In seguito, si è proseguito con il calcolo della media tra i valori ottenuti, per ottenere una soglia di riferimento attraverso cui suddividere i siti in funzione della loro dimensione.

Si sono ottenuti i seguenti valori:

- Per i siti Multibrand la media del n° totale di visite per sito nel periodo di riferimento è di 68585778,4 visite
- Per i siti Monobrand la media del n° totale di visite per sito nel periodo di riferimento è di 18445631,38 visite

Da questi risultati si può comprendere che, per i 289 siti analizzati, il n° di visite per i siti Multibrand nel periodo di riferimento è all'incirca 3 volte e mezzo rispetto al n° di visite per i siti Monobrand.

Dopo aver comparato il n° medio di visite complessive si è ottenuta la suddivisione all'interno di ciascuna categoria tra i siti di grandi dimensioni (sopra il valore medio ottenuto) e quelli classificati come di piccole dimensioni (al di sotto della soglia di riferimento).

I risultati ottenuti per i 226 siti Multibrand sono riportati in Fig. 17 e Fig. 18 in cui si rileva una netta preponderanza per i siti definiti di "piccole dimensioni" rispetto a quelli inseriti nella categoria "Big". Questo è derivato dal fatto che il valore medio è fortemente influenzato dal n° di visite effettuate in siti di dimensioni come Amazon e Alibaba e quindi il valore medio è spinto verso l'alto dividendo i colossi del web dagli altri siti. Questo lo si vede andando ad analizzare la Fig. 21 che riporta su un istogramma i valori rilevati per ciascun sito: i picchi per i siti "Big size" sono nettamente più rilevanti rispetto a tutti gli altri e questo comporta un valore medio delle visite fortemente spinto verso l'alto.

Per i siti Monobrand invece la situazione è un po' più bilanciata in quanto, dei 63 siti individuati, il 22% dei siti è stato inserito nella categoria "Big" e la restante parte nella categoria "Small". Qui la distribuzione è più equilibrata in quanto da come si rileva dalla Fig. 22, l'istogramma riporta picchi meno rilevanti rispetto a quanto riportato per i Multibrand e questo comporta una maggiore omogeneità nella suddivisione tra siti di grandi dimensioni e di piccole dimensioni.

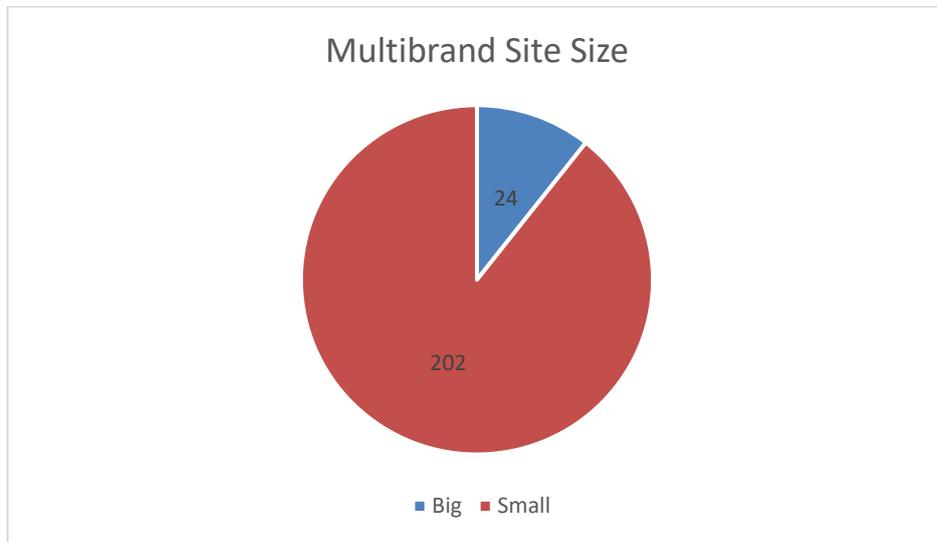


Figura 17: Dimensione siti Multibrand

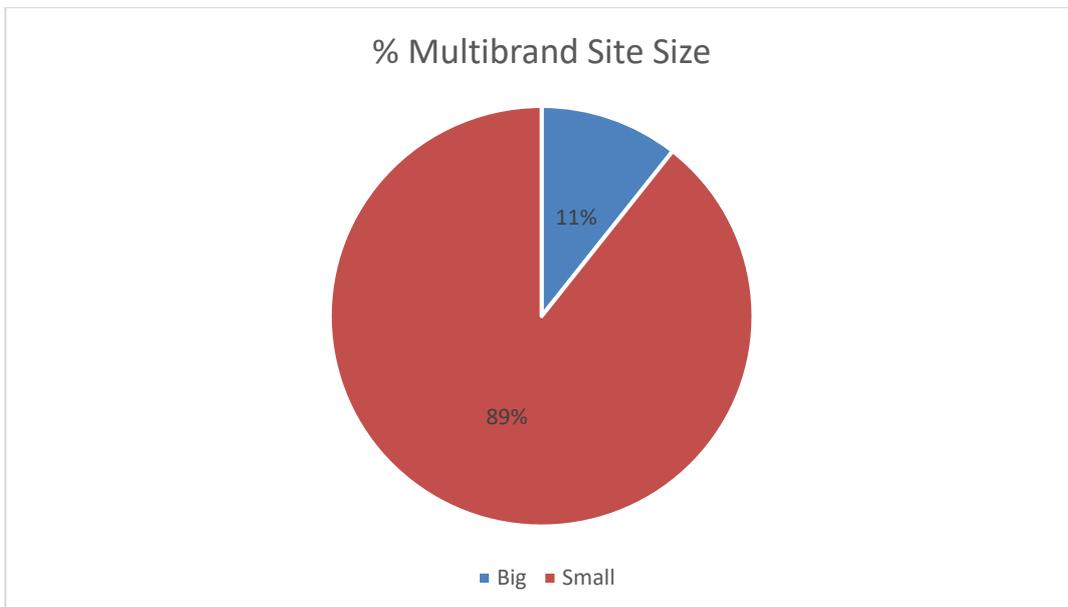


Figura 18: Percentuale di suddivisione in base alla dimensione (Multibrand)

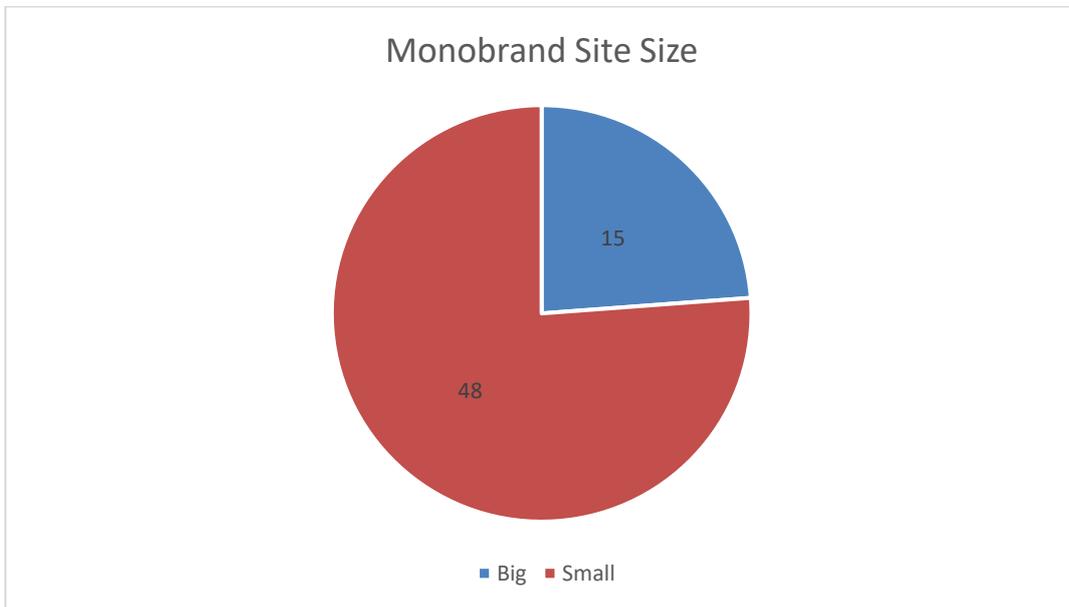


Figura 19: Dimensioni siti Monobrand

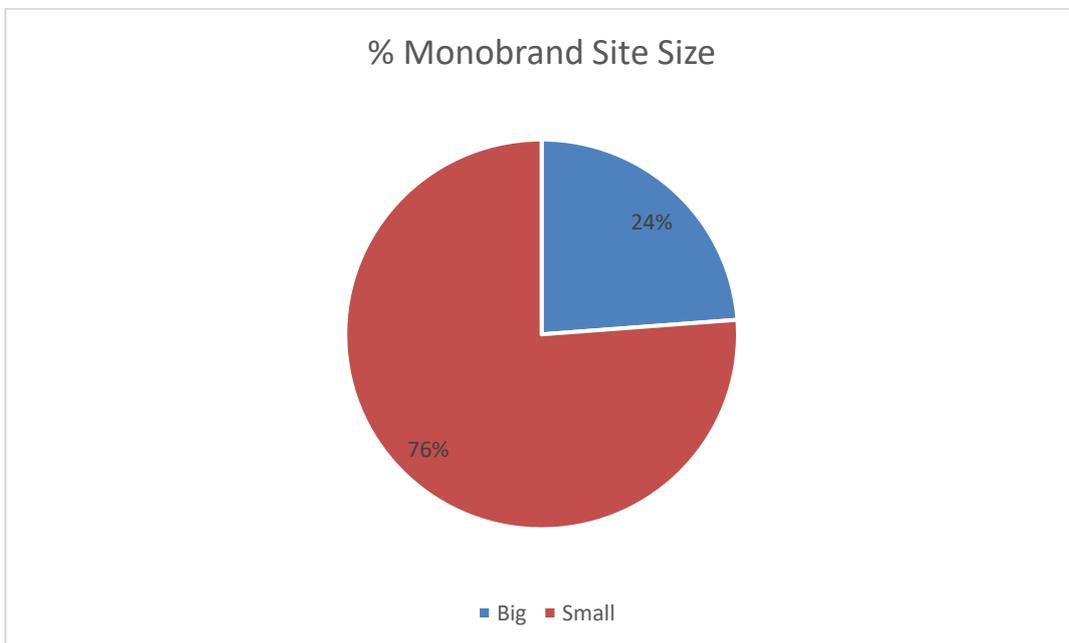


Figura 20: Percentuale di suddivisione in base alla dimensione (Monobrand)

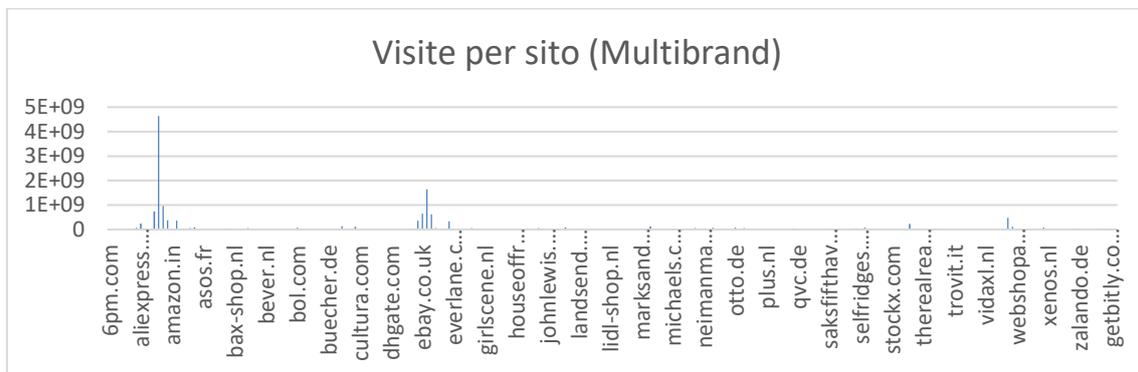


Figura 21: Visite per sito (Multibrand)

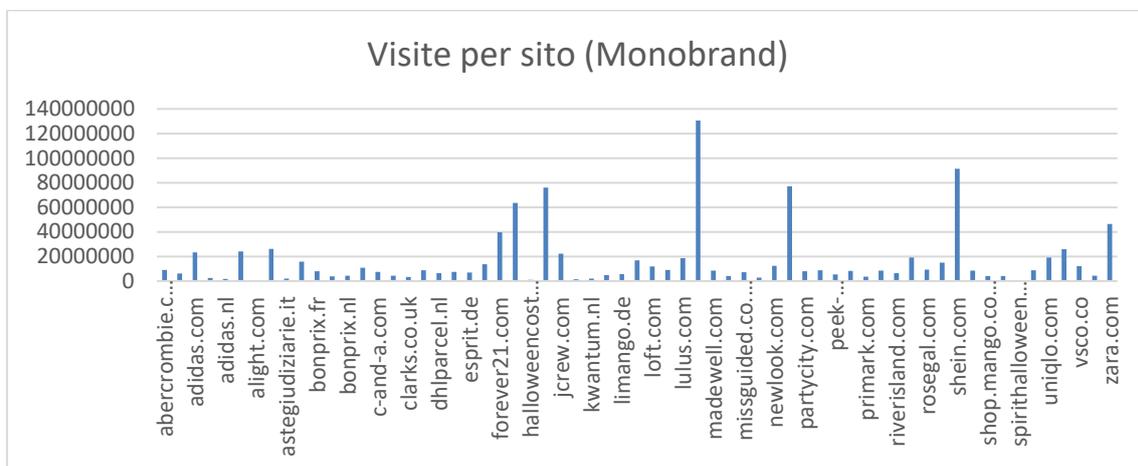


Figura 22: Visite per sito (Monobrand)

Attraverso queste statistiche descrittive sulla base della suddivisione dimensionale dei siti si sono potuti rilevare i seguenti aspetti:

- Nei siti Multibrand il gap tra i colossi e gli altri siti è molto più rilevante rispetto a quanto accade per i siti Monobrand
- Il n° di visite per i siti Multibrand è in media tre volte superiore rispetto a quelle dei Monobrand
- Il mercato per il settore Multibrand è in mano a pochissime realtà e questo comporta ad instaurare una realtà oligopolistica

A valle di questa classificazione si è voluta porre l'attenzione sulla suddivisione geografica delle ricerche effettuate dagli utenti in base ai siti presenti nel dataset e per



questo si è condotta un'analisi per la segmentazione dei siti in funzione del Paese di provenienza della ricerca.

Per fare questo si è seguita la procedura riportata in Fig. 23.

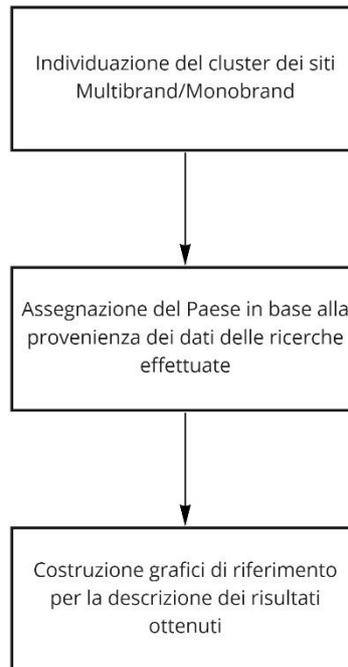


Figura 23: Procedura utilizzate per la segmentazione geografica dei siti

Si è condotta l'analisi sia per i siti Multibrand che per quelli Monobrand, sapendo di avere a disposizione dati provenienti da seguenti Paesi:

- Stati Uniti
- Italia
- UK
- Germania
- Francia
- Olanda

Per avere un quadro complessivo si è deciso di condurre un'analisi aggiuntiva senza separare i siti nelle due macrocategorie, in modo da avere una visione più alta dei dati su base geografica (Fig. 24 e Fig 25.).

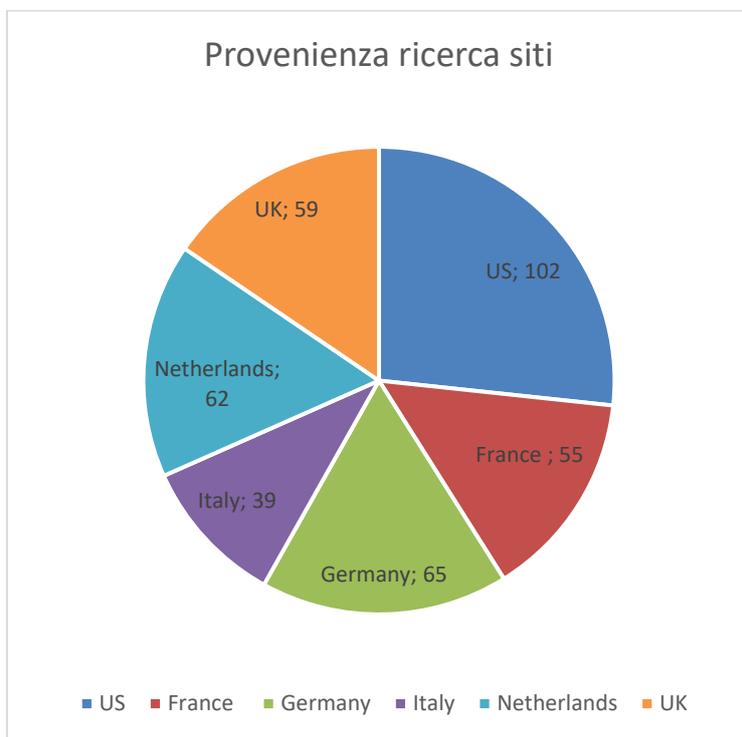


Figura 24: Suddivisione geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento

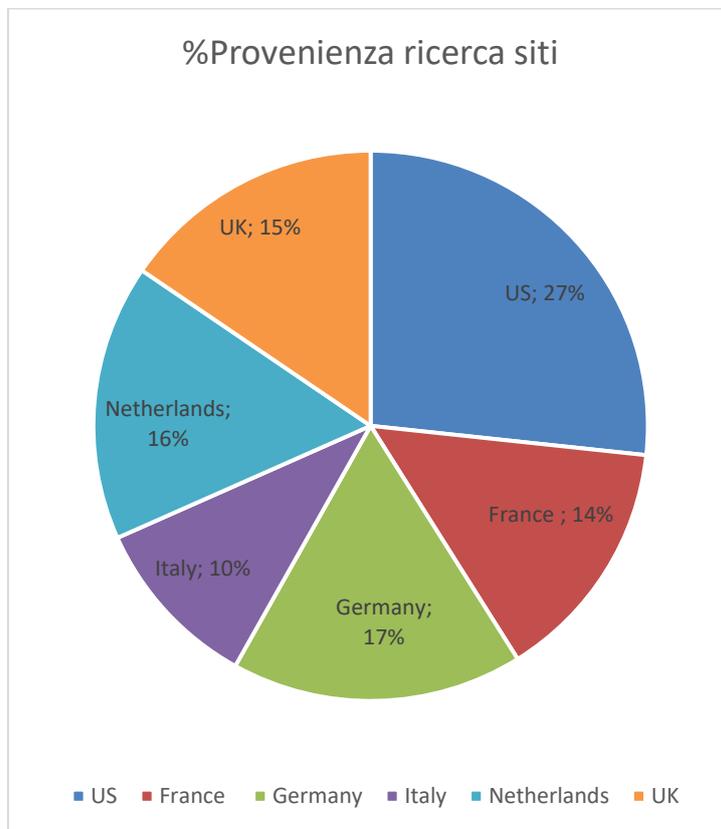


Figura 25: Suddivisione percentuale geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento

Come si può notare le ricerche effettuate nei vari domini di riferimento (siti web) del dataset provengono in maniera pressoché equilibrata da tutti i Paesi di cui si dispongono le informazioni, con una predominanza da parte degli Stati Uniti rispetto ai singoli Paesi europei. Questo permette di ottenere delle valutazioni estremamente significative relativamente alla differenza di traffico tra i Paesi europei (intesi nel loro complesso) e gli Stati Uniti.

In Fig. 26 e in Fig. 27 si riportano le statistiche di suddivisione geografica per i siti Multibrand.

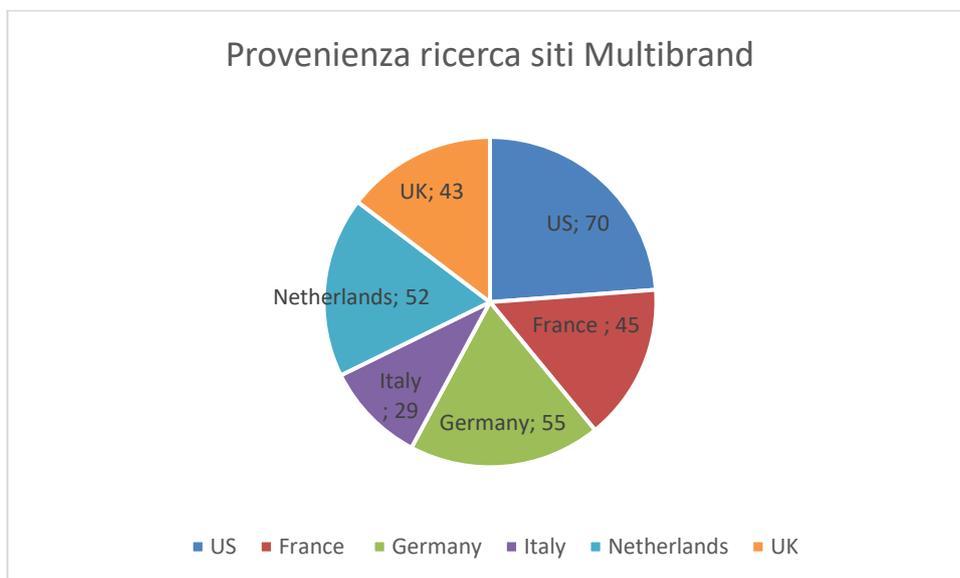


Figura 26: Suddivisione geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento per la categoria Multibrand

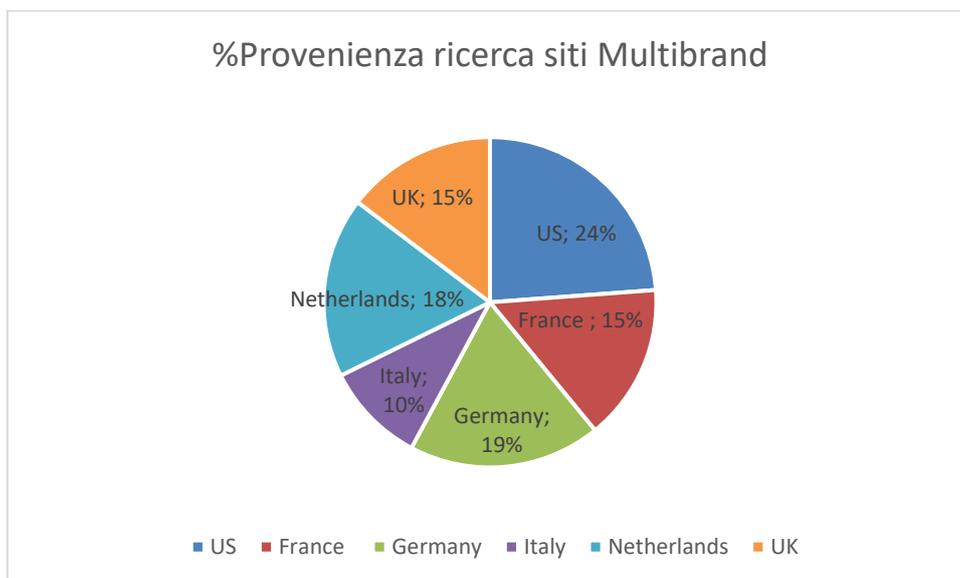


Figura 27: Suddivisione percentuale geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento per i siti Multibrand

Per la categoria dei siti Multibrand si rileva una predominanza degli Stati Uniti, che conferma quanto evidenziato in precedenza, succeduti dalla Germania e dall'Olanda.

In Fig. 28 e in Fig. 29 si riportano invece i risultati dell'analisi condotta sui soli siti Monobrand.

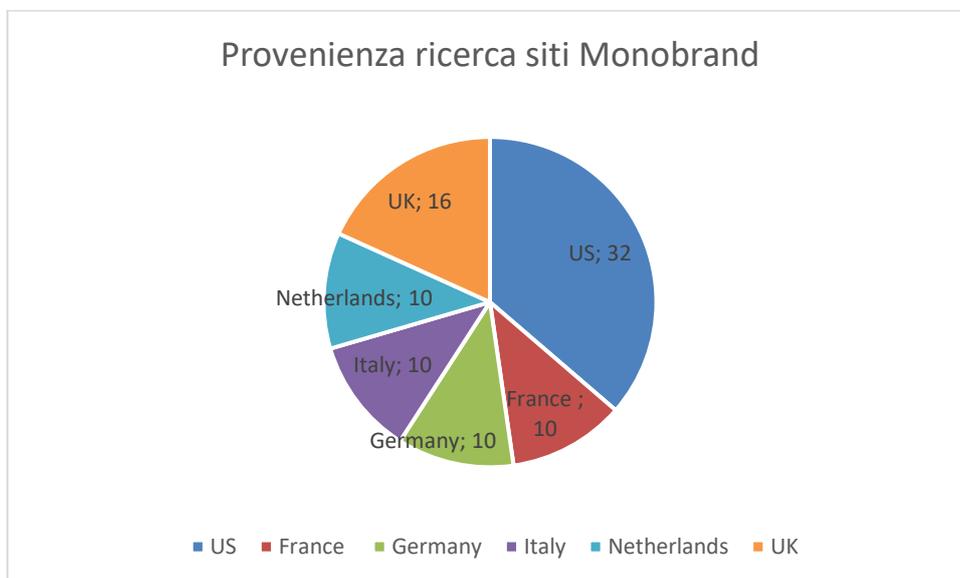


Figura 28: Suddivisione geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento per la categoria Monobrand

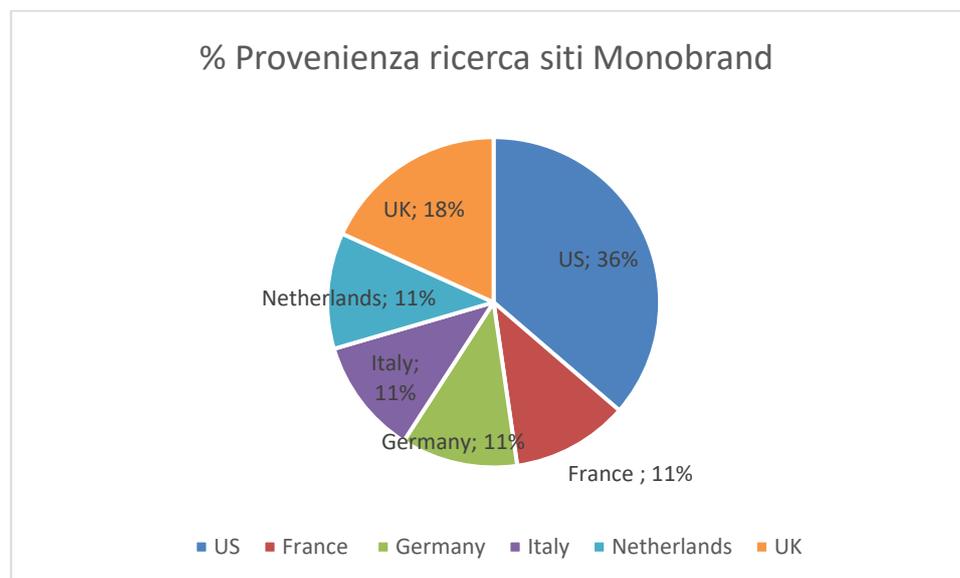


Figura 29: Suddivisione percentuale geografica per Paese di provenienza della ricerca effettuata sul dominio di riferimento per i siti Monobrand

Come in precedenza si evidenzia una predominanza dei dati provenienti dal territorio statunitense, succeduto da UK e dagli altri stati europei.



4.3 Analisi dei trend temporali del traffico online

Dopo aver descritto il dataset a disposizione secondo appunto le modalità riportate nel capitolo precedente ora si vuole proseguire lo studio e l'analisi con la valutazione dei possibili trend temporali di traffico in funzione dell'introduzione del GDPR per i siti Multibrand e Monobrand andando a studiare i dati a disposizione suddividendoli per aree geografiche.

In particolare, si sono analizzati i dati di traffico suddivisi nelle seguenti categorie:

- Siti Multibrand in US
- Siti Multibrand in EU
- Siti Monobrand in US
- Siti Monobrand in EU

Questo ha permesso di studiare le differenze del traffico sia da un punto di vista del segmento industriale che dal punto di vista dell'area geografica.

Le variabili che sono state analizzate, a partire da quelle a disposizione del dataset sono state:

- Il **traffico medio**
- La **durata media della visita**
- Il **numero medio di pagine visitate**
- Il **bounce rate medio**

Il primo passaggio dell'analisi è stato quello di costruire dei dataset di partenza che permettessero di studiare le informazioni sulla base della suddivisione descritta in precedenza, per tale motivo si è costruito un dataset per ogni categoria individuata. Si è iniziato con lo studio delle variabili per i siti Multibrand in Europa e la procedura ha previsto l'individuazione di un trend temporale sulla base del periodo di osservazione dei dati.

Il procedimento seguito è quello individuato in Fig. 30.

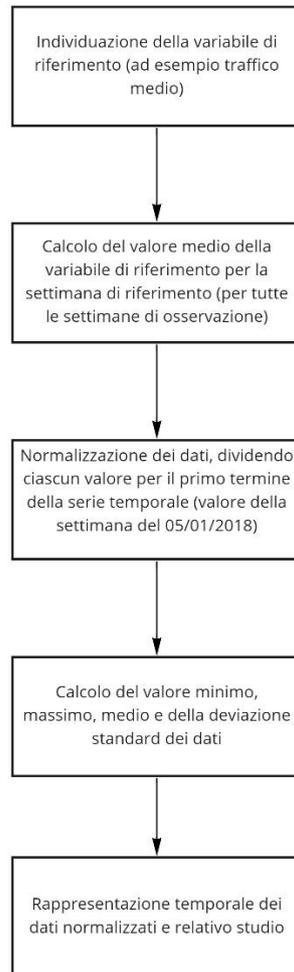


Figura 30: Procedura utilizzata per la valutazione dei trend temporali

Si sono individuati i trend temporali per le variabili sopra descritte per ciascuna delle categorie individuate, effettuando i seguenti confronti:

- Europa vs Stati Uniti per Multibrand
- Europa vs Stati Uniti per Monobrand
- Multibrand vs Monobrand in Europa
- Multibrand vs Monobrand (Europa + Stati Uniti)

Partendo dal primo confronto si riportano i risultati in Fig. 31, 32, 33, 34.

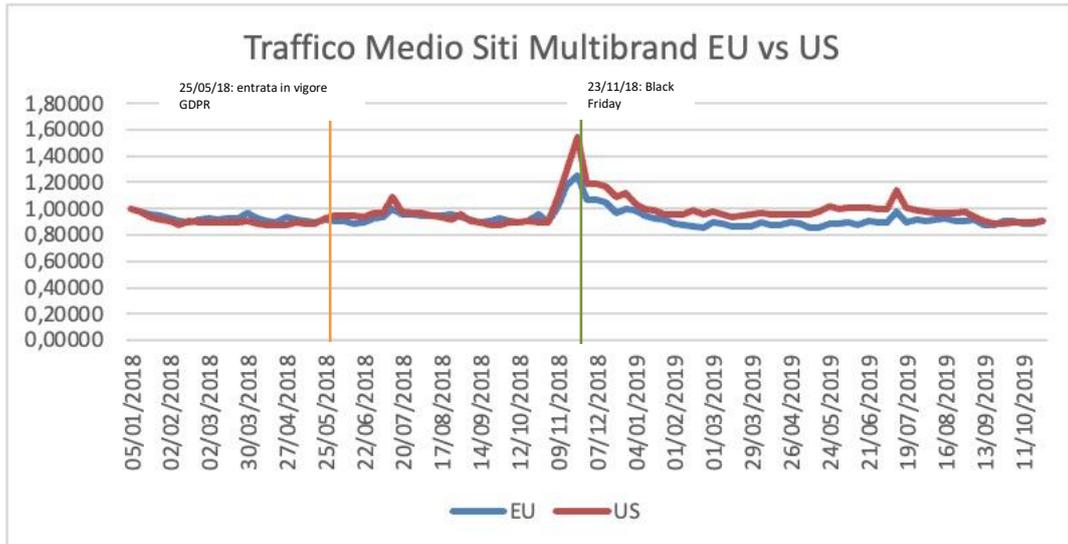


Figura 31: Traffico medio a confronto per siti Multibrand (EU vs US)

L'analisi dei trend temporali sul **traffico medio** riporta due andamenti pressoché analoghi riscontrati tra i siti Multibrand in Europa e negli Stati Uniti: in particolare si evidenzia un picco nel traffico nella settimana del 23/11/18 giorno del Black Friday (giorno in cui le attività commerciali ricorrono a particolare scontistiche per aumentare



il numero di vendite dei propri prodotti), in cui il traffico medio ha riportato valori il 60% più alti rispetto ai dati del periodo precedente.

In seguito a quel picco il traffico è diminuito tornando a valori stazionari sia per gli Stati Uniti che per l'Europa, non riscontrando differenze sostanziali rispetto ai dati registrati precedentemente all'introduzione del GDPR

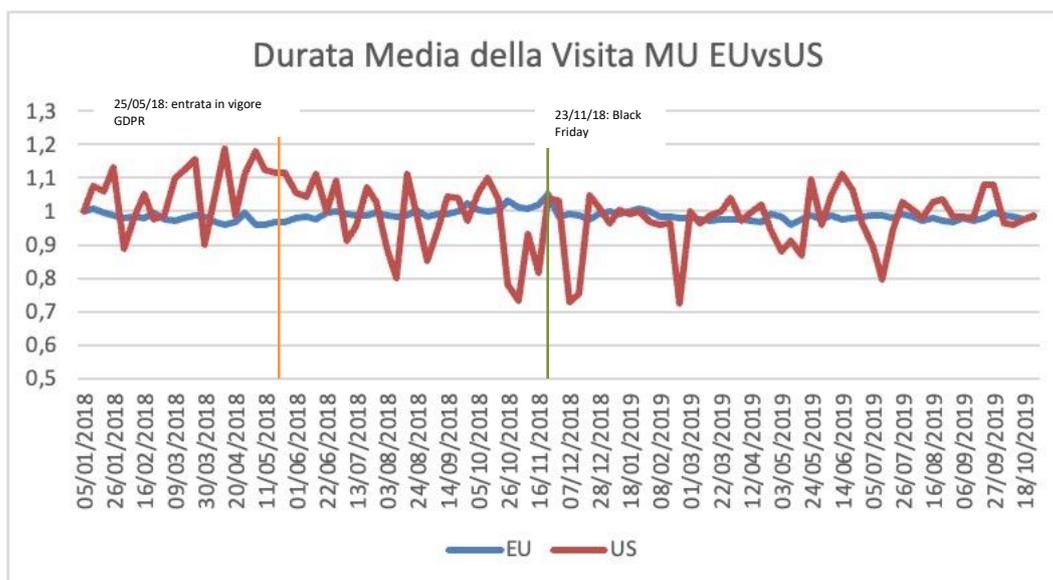


Figura 32: Durata media della visita siti Multibrand (EU vs US)

Per il confronto relativo al trend temporale della durata media della visita nei siti Multibrand si evidenziano i seguenti fattori:

- Il trend per gli Stati Uniti mostra un andamento estremamente incostante rispetto a quanto riscontrato in Europa: negli Stati Uniti si rilevano dei delta del 30-40% (sia positivi che negativi) rispetto al valore registrato nella prima settimana disponibile, questo indica un andamento fortemente variabile nel periodo analizzato.
- Il trend europeo (come per il traffico medio) presenta un picco in corrispondenza della settimana del Black Friday e una conseguente discesa del grafico immediatamente a valle, mentre il trend statunitense presenta il suo massimo a cavallo tra marzo e aprile 2018.
- In Europa non si riscontrano notevoli differenze nella durata media della visita pre e post introduzione del GDPR.

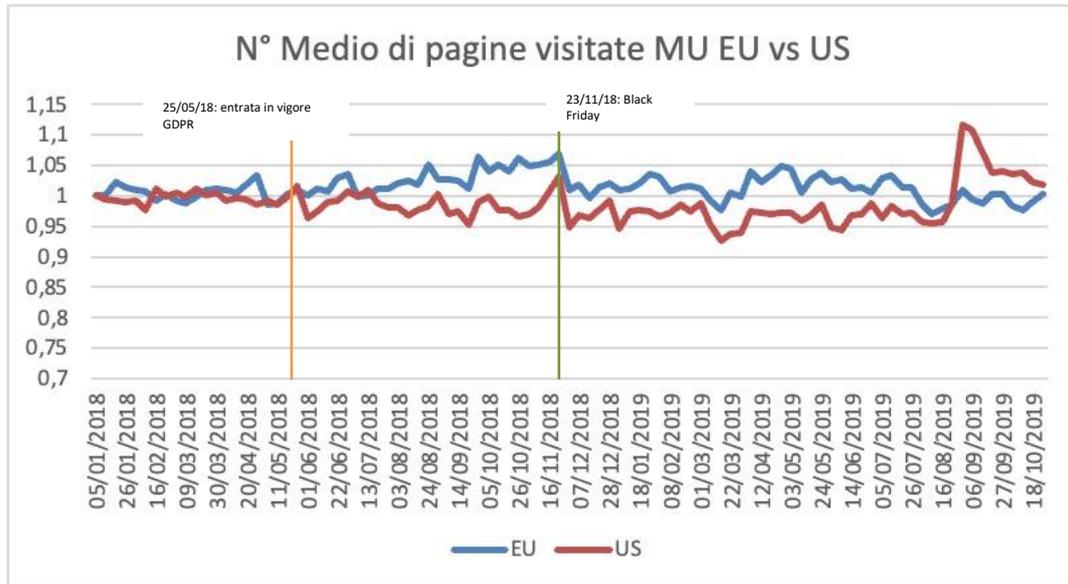


Figura 33: N° medio di pagine visitate di siti Multibrand (EU vs US)



Per quanto riguarda la comparazione del n° medio di pagine visitate tra Europa e Stati Uniti per i siti Multibrand si possono evidenziare i seguenti risultati:

- Il trend europeo presenta un massimo in corrispondenza della settimana del Black Friday (così come evidenziato in precedenza anche in Fig. 32 e 33)
- L'andamento del n° medio di pagine visitate negli Stati Uniti presenta un picco a cavallo tra agosto e settembre del 2019.
- Anche in questo caso il trend europeo decresce a valle della settimana del Black Friday, stazionandosi su valori registrati nelle settimane precedenti, per poi rivedere una crescita in corrispondenza Aprile e Maggio 2019.
- Anche per questa variabile non si evidenziano influenze rilevanti dettate dall'introduzione del GDPR.

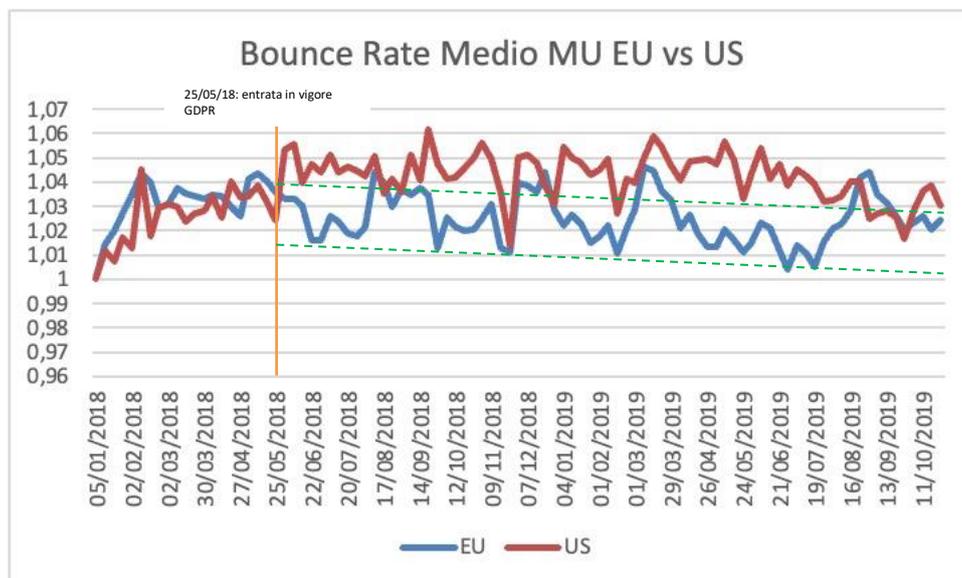


Figura 34: Bounce rate medio per i siti Multibrand (EU vs US)

L'analisi effettuata sui trend relativi agli andamenti del Bounce Rate Medio nei siti Multibrand comparando il traffico US con quello europeo ha portato alle seguenti evidenze:

- Il trend europeo presenta un minimo proprio nella prima settimana rilevata così come il trend statunitense.
- Entrambi i grafici presentano un andamento fortemente variabile evidenziando picchi positivi e negativi.
- Il trend europeo presenta un andamento nel complesso discendente rispetto all'entrata in vigore del GDPR (evidenziato dalle linee tratteggiate in verde nella Fig. 34).



Per il secondo confronto, si sono comparati i trend relativi ai siti Monobrand tra US ed EU, riportati in Fig. 35, 36, 37, 38.

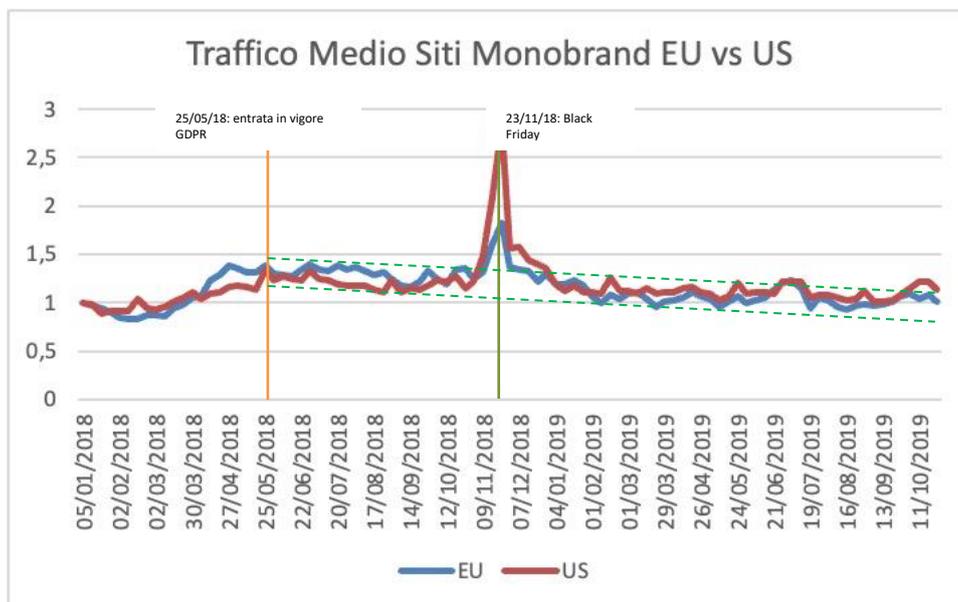


Figura 35: Traffico medio per i siti Monobrand (EU vs US)

Vediamo ora l'analisi effettuata relativa alla comparazione tra i trend temporali del traffico medio nei siti Monobrand in Europa e negli Stati Uniti.

L'analisi effettuata ha portato alle seguenti conclusioni:

- Così come evidenziato per i siti Multibrand, anche il traffico medio, per i siti Monobrand, indipendentemente dal fatto che sia in Europa o negli Stati Uniti, presenta un massimo nella settimana corrispondente al Black Friday (23/11/2018).
- Il trend europeo presenta un andamento decrescente dopo l'introduzione del GDPR (evidenziato dalle linee tratteggiate in verde in Fig. 35), a meno del piccolo riscontrato nella settimana del Black Friday; aspetto riscontrabile anche nel trend statunitense ma con differenze estremamente minori rispetto alla data di introduzione del GDPR.

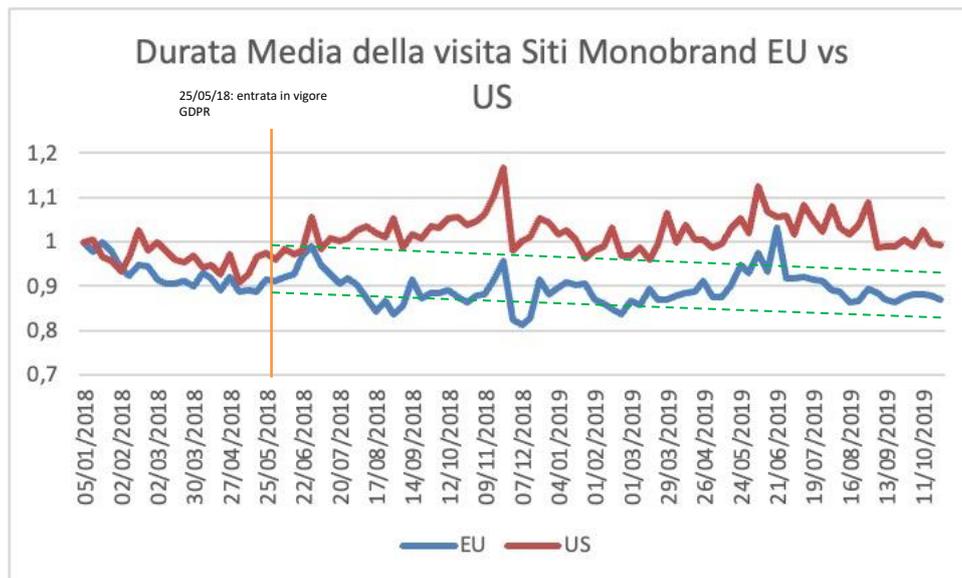


Figura 36: Durata media della visita per i siti Monobrand (EU vs US)

Per quanto riguarda l'analisi effettuata sul trend temporale registrato sulla durata media della visita nei siti Monobrand rispetto all'Europa e agli Stati Uniti, si riportano i seguenti commenti:

- Gli USA presentano un picco sempre in corrispondenza della giornata del Black Friday, mentre per l'Europa il picco si registra nel mese di Giugno 2019.
- L'andamento di entrambi i grafici risulta essere altalenante con picchi valli lungo tutto l'arco temporale, ma il trend europeo presenta un andamento leggermente



decescente dopo l'introduzione del GDPR fino all'ultima settimana registrata, come mostrato dalle linee tratteggiate in verde all'interno della Fig. 36.

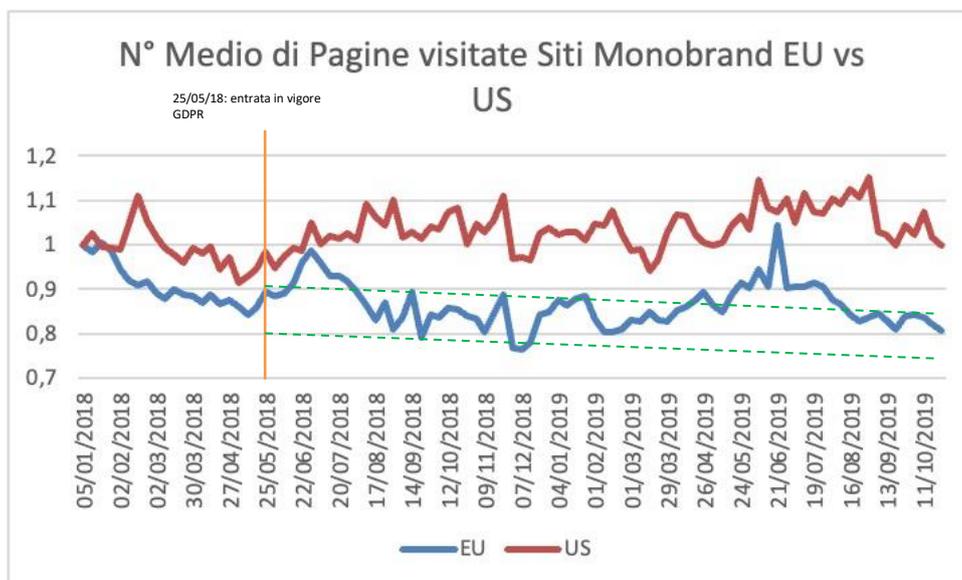


Figura 37: N° medio di pagine visitate per i siti Monobrand (EU vs US)

Per l'analisi sull'andamento del n° di pagine medie visitate nei siti Monobrand per evidenziare le differenze tra Stati Uniti e Europa, si è deciso di riportare le seguenti evidenze:

- Entrambi i trend presentano andamenti variabili lungo l'arco temporale di riferimento senza poter individuare un trend specifico.
- Il trend europeo presenta due picchi: uno a cavallo tra Giugno e Luglio 2018 ed uno a cavallo degli stessi mesi per l'anno successivo
- Il trend statunitense invece due picchi: uno all'inizio dell'estate 2019 e uno in corrispondenza della fine.
- Il trend europeo, a differenza di quello statunitense, presenta un andamento nel complesso discendente rispetto alla data di entrata in vigore del GDPR, come



evidenziate dalle linee tratteggiate in verde riportate in Fig. 37 (a meno dei due picchi precedentemente descritti).

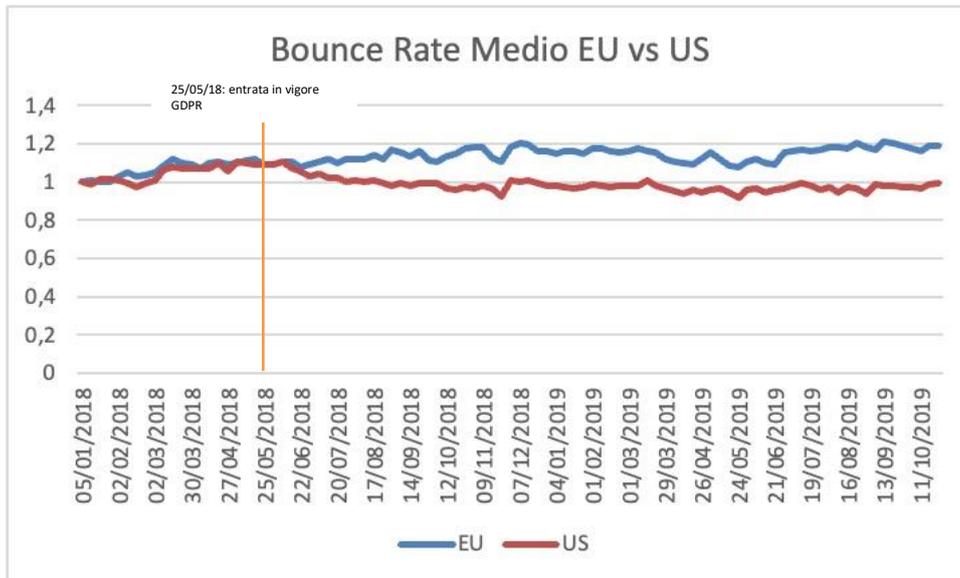


Figura 38: Bounce rate medio per i siti Monobrand (EU vs US)

Per la quarta analisi prevista in questa categoria di confronto si riportano i risultati derivanti dall'analisi effettuata sui trend riscontrati nei siti Monobrand negli Stati Uniti e in Europa:

- In questo caso entrambi i trend presentano una "valle" nel grafico in corrispondenza della settimana del Black Friday.
- Il trend europeo ha risentito dell'introduzione del GDPR in quanto si nota un andamento crescente del valore del Bounce Rate Medio, a testimonianza di una riduzione dell'engagement.
- Il trend statunitense invece sembra presentare un trend leggermente discendente proprio in seguito alla data di entrata in vigore del GDPR, quindi un aumento dell'engagement.

Per il terzo confronto invece si è prevista la comparazione tra i siti Monobrand e quelli Multibrand in EU. I grafici relativi sono riportati in Fig. 39, 40, 41, 42.

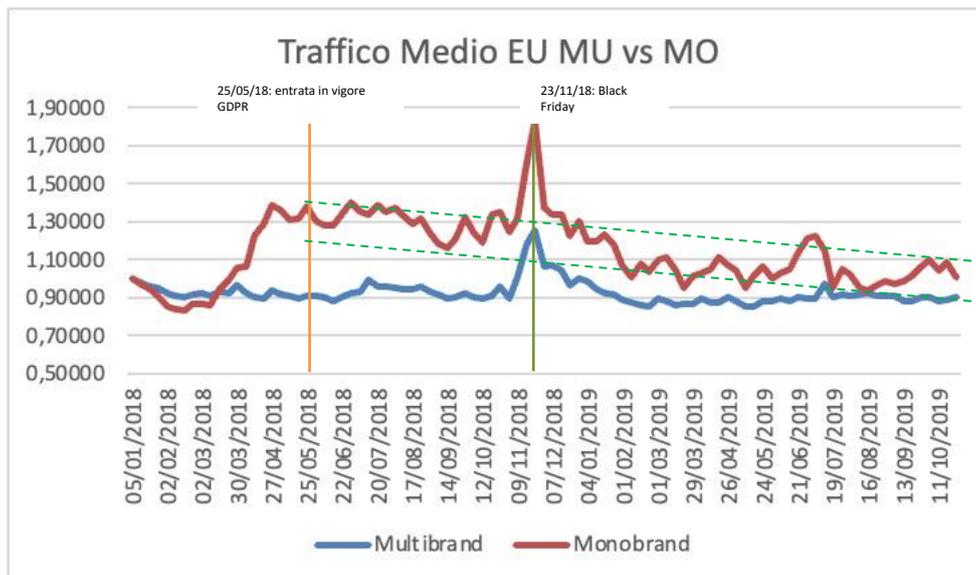


Figura 39: Traffico Medio in EU (Multibrand vs Monobrand)

Per la terza categoria di confronti si è prevista una prima analisi per la comparazione tra i siti Multibrand e i siti Monobrand in Europa, riportando i seguenti risultati:

- Per entrambi i grafici si riporta un picco in corrispondenza della settimana del Black Friday (23/11/18), come riportato dall'etichetta in Fig. 39.
- Si evince un andamento nel complesso discendente per il traffico medio dei siti Monobrand (ancora più evidente rispetto a quanto riscontrato in Fig. 35), a valle dell'entrata in vigore del GDPR, come evidenziato dalle linee tratteggiate verdi in Fig. 39.

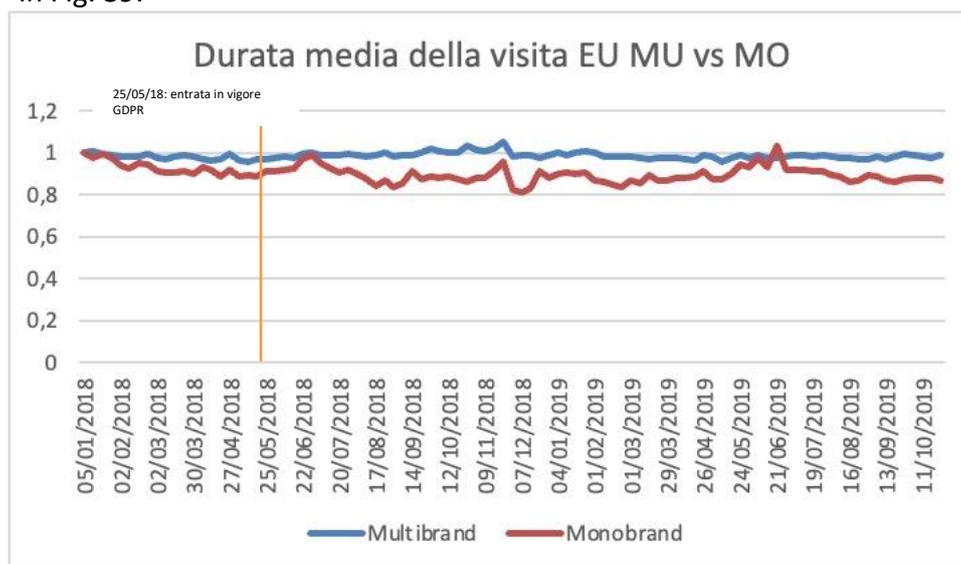


Figura 40: Durata media della visita in EU (Multibrand vs Monobrand)



Per la seconda analisi si riportano i risultati del confronto effettuato tra i siti Multibrand e quelli Monobrand in Europa, relativamente alla durata media della visita:

- Il grafico per i siti Multibrand presenta un massimo in corrispondenza della settimana del Black Friday, aspetto non riscontrato per il grafico dei siti Monobrand.
- Il grafico per i siti Monobrand presenta un andamento molto più variabile rispetto a quello dei siti Multibrand, presentando un andamento leggermente decrescente a valle dell'introduzione del GDPR, aspetto non rilevato per i siti Multibrand.

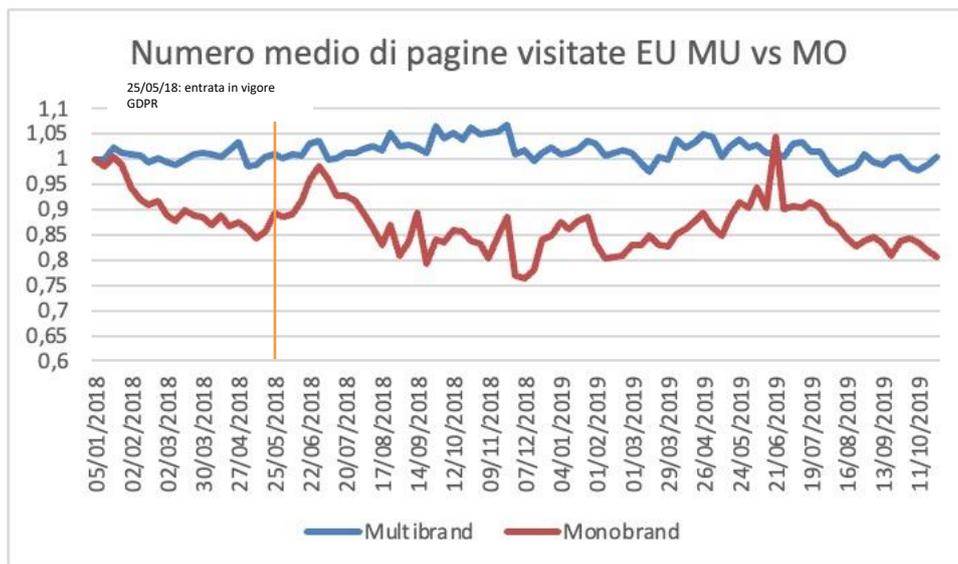


Figura 41: N° medio di pagine visitate in EU (Multibrand vs Monobrand)



Per la terza analisi all'interno del confronto Multibrand vs Monobrand si riportano i risultati relativi al n° medio di pagine visitate in Europa:

- Le differenze sono sostanziali in quanto i siti Multibrand presentano un picco in corrispondenza della settimana del Black Friday, a differenza dei siti Monobrand che presentano un picco tra Giugno e Luglio 2019.
- Il trend dei siti Monobrand è complessivamente decrescente a meno dei due picchi registrati, mentre il grafico dei siti Multibrand sembra non aver risentito dell'entrata in vigore del GDPR.

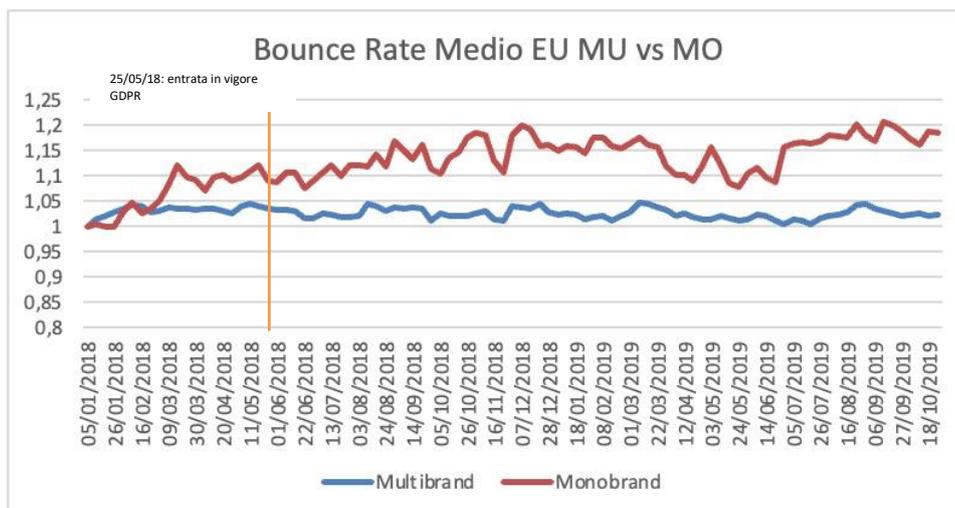


Figura 42: Bounce rate medio in EU (Multibrand vs Monobrand)

Per l'ultima analisi di quest'ultimo confronto si riportano i grafici relativi al trend del Bounce Rate Medio dei siti Monobrand e di quelli Multibrand, riscontrando le seguenti differenze:

- Il grafico dei siti Monobrand presenta un trend ascendente a partire dalla prima settimana di osservazione, mentre l'andamento per i siti Multibrand è pressoché costante; questo determina una riduzione dell'engagement per i siti Monobrand, a differenza di un andamento pressoché costante per i siti Multibrand.
- Per i siti Monobrand si presenta un picco negativo in corrispondenza del Black Friday così come per quelli Multibrand.

L'ultimo confronto invece ha riguardato in termini assoluti una comparazione tra i siti Multibrand e quelli Monobrand. A tal fine si sono comparati i differenziali dei dati normalizzati di US e EU, tra i siti Multibrand e quelli Monobrand:

- Si è calcolato il differenziale di ciascuna variabile normalizzata all'interno della categoria (prima Multibrand e poi Monobrand) tra US e EU.
- Si è graficato il trend dei differenziali (i differenziali sono stati sempre calcolati andando a sottrarre dal valore normalizzato US quello EU), come riportato in



Fig. 43, andando a delineare le differenze tra l'andamento del grafico dei siti Multibrand e di quelli Monobrand.

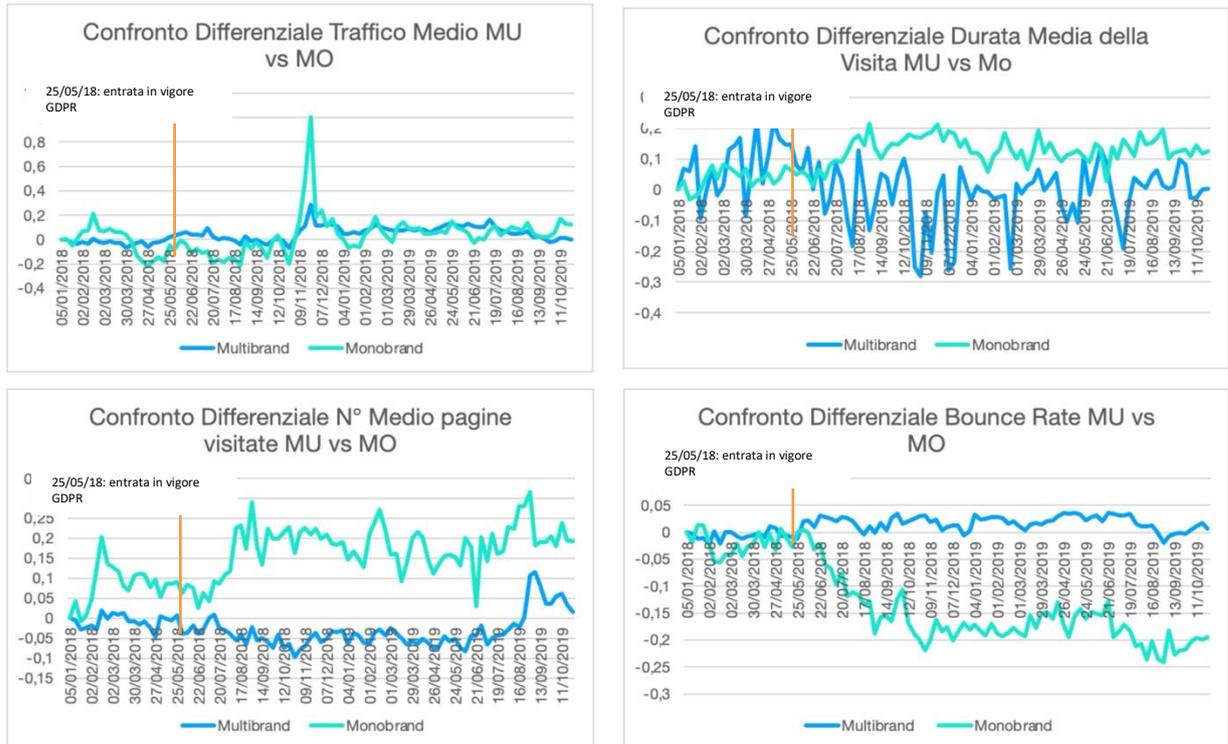


Figura 43: Confronto tra i differenziali US e EU delle variabili individuate tra siti Multibrand e siti Monobrand

Facendo questo confronto è stato facile comprendere i seguenti aspetti:

- Il traffico medio è stato influenzato maggiormente nel Monobrand rispetto al Multibrand, come si può vedere nel primo pannello in alto a sinistra in quanto si è passati da valori differenziali negativi pre GDPR per arrivare a valori differenziali positivi mentre i differenziali per il Multibrand si sono mantenuti pressochè su valori positivi, ciò vuol dire che è aumentato il delta tra i valori del traffico statunitense con quelli europei per i siti Monobrand e non si è registrata la stessa variazione per i siti Multibrand.
- I trend invece relativi ai differenziali della durata media di una visita mostrano una netta inversione tra la curva verde acqua e quella azzurro scuro dopo l'introduzione del GDPR; questo implica un aumento del differenziale per i siti Monobrand e quindi una conseguente riduzione del tempo che un utente dedica alla propria visita in Europa rispetto a quanto emerso negli Stati Uniti.
- Per quanto riguarda il N° medio di pagine visitate il discorso non cambia in quanto il delta tra la curva Monobrand e quella Multibrand aumenta a valle



**Politecnico
di Torino**

dell'entrata in vigore del GDPR ed è chiaro anche qui l'incremento del delta tra Stati Uniti ed Europa verificato nella categoria Monobrand rispetto a quanto emerso nei siti Multibrand.

- Il grafico invece relativo al Bounce Rate riporta una netta riduzione dell'engagement a valle dell'introduzione del GDPR per i siti Monobrand, aspetto che non riscontrato in alcun modo nei siti Multibrand, dove il valore differenziale sembra avere un trend pressoché costante.



4.4 Conclusioni sulla base dell'analisi dei trend temporali delle variabili individuate

Sulla base di quanto emerso dalle analisi grafiche e statistiche effettuate sui trend temporali relativi al traffico medio, durata media della visita, n° medio di pagine visitate e bounce rate medio, si può concludere che:

- Esiste una differenza sostanziale tra il traffico online negli Stati Uniti ed in Europa sia per i siti Multibrand che per i Monobrand, in particolare:
 - Il trend della durata media di una visita è fortemente differente tra Europa e Stati Uniti sia per Monobrand che per Multibrand (trend discendente per i Monobrand in Europa a valle dell'entrata in vigore del GDPR)
 - I grafici per il traffico medio presentano invece delle analogie (picco in corrispondenza del Black Friday) seppur per i siti Monobrand in EU si presenta un andamento decrescente
 - Per il n° medio di pagine visitate si verifica un trend discendente per i siti Monobrand in Europa a differenza di quanto accade negli Stati Uniti, a valle dell'introduzione del GDPR
 - Il Bounce Rate invece presenta in ogni caso un picco negativo in corrispondenza del Black Friday, in quanto, per sua definizione, rileva la percentuale di visitatori che lascia il sito web dopo aver visitato una sola pagina ed è chiaro che durante il Black Friday l'utente tende ad effettuare una ricerca più prolungata e dettagliata
- Per quanto riguarda invece la differenza in Europa tra siti Monobrand e Multibrand, i risultati rispecchiano e confermano lo studio descritto nella Sezione 3.2.3 in quanto per i siti Monobrand si rileva un trend discendente per tutte le statistiche, ad eccezione del Bounce Rate, e questi, rispetto a quelli Multibrand, che non riportano questo andamento in riduzione, sono di dimensioni molto minori in termini di traffico e conferma quanto rilevato dal precedente studio.
- In termini di differenziali, che è l'analisi maggiormente significativa in quanto valuta in termini assoluti le differenze che incorrono tra i siti Monobrand e quelli Multibrand, è emersa la grande influenza che ha avuto l'entrata in vigore del GDPR nelle variabili relative ai siti Monobrand a differenza di quanto accaduto ai dati del traffico dei siti Multibrand che non sembrano aver risentito in particolar modo l'ingresso del GDPR

In conclusione, è chiaro come il GDPR abbia influenzato il traffico europeo andando a causare una forte riduzione delle variabili studiate per i siti Monobrand rispetto a



**Politecnico
di Torino**

quanto accaduto per i siti Multibrand, a conferma dello studio effettuato in funzione delle dimensioni dei siti, descritto nelle sezioni precedenti.



5. Analisi del modello di regressione

In questo capitolo, al fine di valutare la significatività dei risultati descritti nella Sez. 4, si sono riportati gli studi effettuati attraverso l'implementazione del modello di regressione su STATA per l'analisi degli effetti dell'introduzione del GDPR e della dislocazione geografica dei domini di ricerca (EU vs US), in funzione delle categorie di E-commerce Multibrand e Monobrand.

È stata implementata la seguente equazione:

$$y_{D,t,c} = \beta_0 + \beta_1 post_t + \beta_2 treat_c + \beta_3(post \times treat)_{t,c} + \beta_4(mo \times post \times treat) \varepsilon$$

Dove la y indica la variabile in esame tra quelle di interesse, ovvero:

- Traffico (tramite trasformazione logaritmica);
- Durata media delle visite (tramite trasformazione logaritmica);
- Numero medio di pagine visitate (tramite trasformazione logaritmica);
- Bounce rate

Mentre per le altre variabili:

- **post** assume valore 1 se il dato in esame è successivo al GDPR (funzione del tempo t);
- **treat** assume valore 1 se il paese di provenienza della visita appartiene all'Europa (Italia, UK, Germania, Francia e Olanda) e 0 se il dato di provenienza è dall'US (funzione del country c);
- **post x treat** rappresenta l'iterazione tra le precedenti, perciò assume valore 1 soltanto se la visita è europea e successiva al GDPR;
- **mo x post x treat**: rappresenta l'interazione tra le variabili precedenti, e assume il valore di 1 quando la visita è stata effettuata dopo il GDPR, in Europa e si riferisce ad un dominio della categoria Monobrand
- ε rappresenta il termine di errore.

Per ciascuna variabile (traffico, durata media della visita, numero medio delle pagine visitate e bounce rate) sono state implementati quattro modelli di regressione per determinare la significatività delle variazioni comportate sia dalla provenienza delle ricerche che dal GDPR:

- **Interazione base**



- **Specificazione 1:** considera gli effetti fissi sulla variabile dominio
- **Specificazione 2:** valuta gli effetti fissi su dominio e settimane
- **Specificazione 3:** valuta gli effetti fissi su dominio, settimane e provenienza del sito di ricerca (country)
- **Specificazione 4:** per valutare se ci siano stati degli effetti eterogenei tra la categoria Multibrand e Monobrand derivanti dall'introduzione del GDPR

In Tab. 7 si riportano i risultati relativi alla variabile **Traffico**, frutto del modello di regressione implementato su STATA, secondo i criteri e le variabili precedentemente descritte.

Tabella 7: Analisi di significatività della variazione del traffico

VARIABLES	(1) log_traffic	(2) log_traffic	(3) log_traffic	(4) log_traffic	(5) log_traffic
postxtreat	-0.114** (0.0450)	-0.119*** (0.0451)	-0.119*** (0.0452)	-0.119*** (0.0452)	-0.157*** (0.0512)
moxpostxtreat					0.191 (0.173)
treat	-1.142*** (0.128)	-1.843*** (0.293)	-1.842*** (0.294)		
post	0.155*** (0.0374)	0.156*** (0.0379)			
Constant	13.87*** (0.125)	14.39*** (0.216)	14.51*** (0.216)	12.74*** (0.251)	12.73*** (0.253)
Observations	35,988	35,988	35,988	35,988	35,988
R-squared	0.170	0.703	0.706	0.765	0.765

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

I risultati ottenuti hanno portato alle seguenti conclusioni per quanto riguarda la variabile traffico:

- Risulta esserci stata una riduzione statisticamente significativa per i valori di **Traffico** in correlazione all'entrata del GDPR (colonna 1, 2, 3, 4)
- Non risulta essere statisticamente significativa invece la variazione legata alle differenze riscontrate nei valori di Traffico tra la categoria Monobrand con quella Multibrand (evidenziata in colonna 5)



**Politecnico
di Torino**

Quindi sono stati riscontrate delle differenze legate all'introduzione del GDPR così come evidenziato nella Sez. 4 (in particolare per la categoria Monobrand) trasversalmente alle categorie Multibrand e Monobrand, ma non è stata rilevata una variazione statisticamente significativa per quanto riguarda i valori registrati tra la categoria Monobrand e quella Multibrand

In Tab. 8 si riportano i risultati relativi alla variabile **Durata media della visita**, frutto del modello di regressione implementato su STATA, secondo i criteri e le variabili precedentemente descritte.



Tabella 8: Analisi della significatività della variazione della durata media della visite

VARIABLES	(1) log_visit_duration	(2) log_visit_duration	(3) log_visit_duration	(4) log_visit_duration	(5) log_visit_duration
postxtreat	-0.0194 (0.0167)	-0.0196 (0.0161)	-0.0196 (0.0161)	-0.0196 (0.0161)	-0.0135 (0.0170)
moxpostxtreat					-0.0302 (0.0550)
treat	-0.0742 (0.0547)	-0.0645 (0.0786)	-0.0646 (0.0787)		
post	0.0151 (0.0142)	0.0177 (0.0138)			
Constant	5.866*** (0.0426)	5.857*** (0.0582)	5.867*** (0.0579)	5.821*** (0.0370)	5.822*** (0.0380)
Observations	35,988	35,988	35,988	35,988	35,988
R-squared	0.006	0.879	0.880	0.881	0.881

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

I risultati ottenuti hanno portato alle seguenti conclusioni per quanto riguarda la variabile durata media della visita:

- Dalla colonne 1,2,3 e 4 non risultano esserci variazioni statisticamente significative derivanti dall'introduzione del GDPR o dalla provenienza geografica della ricerca effettuata
- Dalla colonna 5 invece si evince che non ci sono differenze statisticamente significative tra la categoria Multibrand e quella Monobrand

I risultati ottenuti dall'implementazione dei modelli di regressione, quindi, hanno comportato una non significatività della variazione della variabile in esame a differenza di quanto riscontrato nei trend temporali analizzati nella Sez. 4 che avevano evidenziato un decremento della durata media della visita, soprattutto per la categoria Monobrand. Queste discrepanze possono essere giustificate dalle assunzioni, aggregazioni e normalizzazioni che sono state effettuate per i calcoli in Excel.



**Politecnico
di Torino**

In Tab. 9 si riportano i risultati relativi alla variabile **Numero medio di pagine visitate**, frutto del modello di regressione implementato su STATA, secondo i criteri e le variabili precedentemente descritte.



Tabella 9: Analisi della significatività della variazione del numero medio di pagine visitate

VARIABLES	(1) log_pages_visited	(2) log_pages_visited	(3) log_pages_visited	(4) log_pages_visited	(5) log_pages_visited
postxtreat	-0.0283 (0.0202)	-0.0289 (0.0192)	-0.0288 (0.0192)	-0.0288 (0.0192)	-0.0206 (0.0198)
moxpostxtreat					-0.0409 (0.0554)
treat	-0.00346 (0.0661)	-0.0495 (0.0761)	-0.0497 (0.0762)		
post	0.0125 (0.0170)	0.0147 (0.0162)			
Constant	2.088*** (0.0559)	2.120*** (0.0576)	2.122*** (0.0554)	2.063*** (0.0367)	2.065*** (0.0380)
Observations	35,988	35,988	35,988	35,988	35,988
R-squared	0.000	0.869	0.869	0.870	0.870

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

I risultati ottenuti hanno portato alle seguenti conclusioni per quanto riguarda la variabile numero medio di pagine visitate:

- Dalla colonne 1,2,3 e 4 non risultano esserci variazioni statisticamente significative derivanti dall'introduzione del GDPR o dalla provenienza geografica della ricerca effettuata
- Dalla colonna 5 invece non risultano esserci delle differenze statisticamente significative tra la categoria Multibrand e quella Monobrand

I risultati ottenuti dall'implementazione dei modelli di regressione, quindi, hanno rilevato una non significatività nella variazione del numero medio di pagine visitate a differenza di quanto riscontrato nei trend temporali analizzati nella Sez. 4 che avevano evidenziato un decremento nel valore della variabile in esame, sia per la categoria Monobrand che Multibrand. Queste discrepanze, come per il caso precedente, possono essere giustificate dalle assunzioni, aggregazioni e normalizzazioni che sono state effettuate per i calcoli in Excel.



**Politecnico
di Torino**

In Tab. 10 si riportano i risultati relativi alla variabile **Bounce Rate**, frutto del modello di regressione implementato su STATA, secondo i criteri e le variabili precedentemente descritte.



Tabella 10: Analisi della significatività della variazione del Bounce Rate

VARIABLES	(1) bounce rate	(2) bounce rate	(3) bounce rate	(4) bounce rate	(5) bounce rate
postxtreat	0.00480 (0.00574)	0.00565 (0.00542)	0.00562 (0.00542)	0.00561 (0.00542)	0.00300 (0.00572)
moxpostxtreat					0.0130 (0.0155)
treat	-0.00221 (0.0154)	0.0138 (0.0235)	0.0138 (0.0235)		
post	-0.00211 (0.00511)	-0.00283 (0.00484)			
Constant	0.320*** (0.0127)	0.308*** (0.0175)	0.305*** (0.0173)	0.322*** (0.0112)	0.322*** (0.0115)
Observations	35,988	35,988	35,988	35,988	35,988
R-squared	0.000	0.905	0.906	0.906	0.907

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

I risultati ottenuti hanno portato alle seguenti conclusioni per quanto riguarda la variabile bounce rate:

- Dalla colonne 1,2,3 e 4 non risultano esserci variazioni statisticamente significative derivanti dall'introduzione del GDPR o dalla provenienza geografica della ricerca effettuata
- Dalla colonna 5 invece non risultano esserci, come per le due casistiche precedenti, delle differenze statisticamente significative tra la categoria Multibrand e quella Monobrand

I risultati ottenuti dall'implementazione dei modelli di regressione, quindi, hanno rilevato una non significatività nella variazione dell'engagement a differenza di quanto riscontrato nei trend temporali analizzati nella Sez. 4 che avevano evidenziato un aumento del valore della variabile in esame, soprattutto nella categoria Monobrand. Queste discrepanze, come per le due casistiche precedenti, possono essere giustificate dalle assunzioni, aggregazioni e normalizzazioni che sono state effettuate per i calcoli in Excel.



6. Conclusioni e riflessioni finali

In conclusione, questa tesi ha prodotto dei risultati estremamente significativi per quanto riguarda, in fase iniziale, i trend che fino ad ora sono stati trattati dalla letteratura scientifica e relativamente all'introduzione del GDPR per quanto riguarda le variazioni nei valori di traffico, durata media della visita, numero medio di pagine visitate e bounce rate (inverso dell'engagement).

Per il primo tema, si riporta che:

- Il tema del GDPR è un aspetto di estremo interesse per la letteratura scientifica sia da un punto di vista dell'applicazione sui mercati ma allo stesso tempo per il risvolto che questa normativa avrà sulla tutela della privacy delle persone
- Gli effetti del GDPR sono visibili, ma ancora poco definiti in molti settori, a causa dell'orizzonte temporale limitato che è intercorso dalla data di entrata in vigore
- Molti settori dovranno essere ancora impattati dalla normativa del GDPR, quali ad esempio quello medico, che richiederà ulteriori approfondimenti nei prossimi anni, soprattutto a valle del periodo pandemico che ha caratterizzato l'ultimo biennio nel quale il trattamento dei dati è stato uno dei temi più rilevanti tra quelli affrontati nell'ultimo periodo

Relativamente ai risultati ottenuti sull'analisi dei trend dei dati del traffico si rimanda a quanto già descritto nella Sez. 2.4; riassumendo:

- Per la categoria Monobrand le variabili in esame sembrano aver risentito dell'introduzione del GDPR in misura maggiore rispetto a quanto accaduto per la categoria Multibrand che sembra essere meno impattata dall'introduzione del GDPR
- Inoltre, sono emerse differenze anche nel confronto del traffico tra aree geografiche anche all'interno delle singole categorie, in particolare per la categoria Monobrand

I risultati dei modelli di regressione per valutare la significatività statistica delle variazioni riscontrate nelle variabili in esame hanno fatto emergere i seguenti risultati, come descritto nella Sez 5:

- Per la variabile traffico le variazioni riscontrate risultano essere statisticamente significative ad eccezione dell'eterogeneità degli effetti sulle categorie Multibrand e Monobrand della quale non risulta esserci rilevanza statistica
- Per le restanti tre variabili:
 - Durata media della visita
 - Numero medio di pagine visitate
 - Bounce rate



**Politecnico
di Torino**

Non risultano esserci variazioni statistiche significativamente rilevanti sia in termini assoluti che all'interno della categoria Multibrand e Monobrand

Come specificato in precedenza, le differenze ottenute tra l'analisi dei trend temporali e le analisi di regressione possono essere portate dalle approssimazioni e aggregazioni che sono state previste per il calcolo in Excel per i primi studi effettuati.

Per concludere, quindi è possibile affermare che il GDPR ha avuto un impatto sull'E-commerce sia all'interno della categoria Monobrand e Multibrand (soprattutto per la variabile traffico), ma non risultano essere statisticamente significative le variazioni riscontrate per le altre variabili in esame, nonostante siano state evidenziate delle modifiche nei trend temporali a valle dell'introduzione del GDPR.

Sarà opportuno, nel corso dei prossimi anni, studiare la variazioni delle suddette variabili rispetto ad un orizzonte temporale più lungo in modo da avere maggiori evidenze su eventuali variazioni dovute appunto all'entrata in vigore del GDPR del 25/05/2018.



Appendice A

Siti Multibrand					
6pm.com	blancheporte.fr	etsy.com	lidl.fr	qualescegliere.it	trustedshops.com
abebooks.com	blokker.nl	everlane.com	macrolibrarsi.it	qvc.com	unitedwardrobe.com
aboutyou.de	bloomingdales.com	expertentesten.de	mandmdirect.com	qvc.de	urbanoutfitters.com
aboutyou.nl	bmstores.co.uk	fashionnova.com	manners.nl	real.de	venteprivee.com
action.com	bol.com	femmeactuelle-news.fr	mariefrance.fr	rebuy.de	vergleich.org
aldi.nl	booklooker.de	fnac.com	marksandspencer.com	reclamefolder.nl	very.co.uk
alibaba.com	boots.com	galerieslafayette.com	markt.de	reservix.de	vidaxl.nl
aliexpress.com	brandalley.fr	gemo.fr	marktplaats.nl	revolve.com	vide-greniers.org
allegro.pl	brands4friends.de	girlscene.nl	matalan.co.uk	ruedcommerce.fr	vinted.fr
amazon.ca	breuninger.com	globalsources.com	mediamarkt.nl	saksfifthavenue.com	vivaticket.it
amazon.co.uk	bricodpot.fr	halfords.com	medimops.de	saldiprivati.com	volantinofacile.it
amazon.com	buecher.de	hema.nl	mercari.com	samsclub.com	walmart.com
amazon.de	carrefour.fr	hobbycraft.co.uk	michaels.com	sarenza.com	wayfair.com
amazon.fr	catawiki.nl	home24.de	missyusa.com	schnaepchenfuchs.com	webshopapp.com
amazon.in	cdiscout.com	hottopic.com	mobiel.nl	scholastic.com	wehkamp.nl
amazon.it	coolblue.nl	houseoffraser.co.uk	myhermes.co.uk	sears.com	weltbild.de
anthropologie.com	costco.co.uk	hse24.de	mysupermarket.co.uk	selfridges.com	westelm.com
appliancesdirect.co.uk	costco.com	hsn.com	narvar.com	shop.nordstrom.com	wikibuy.com
argos.co.uk	cultura.com	ibs.it	nectar.com	showroomprive.com	wilko.com
asos.com	dealsea.com	intermarche.com	neimanmarcus.com	shpock.com	wish.com



asos.de	debenhams.com	jcpenny.com	next.co.uk	sovendus.com	xenos.nl
asos.fr	debijenkorf.nl	jdsports.nl	nordstrom.com	spartoo.com	yoox.com
auchan.fr	decathlon.de	johnlewis.com	nordstromrack.com	stitchfix.com	zalando-lounge.de
audible.de	deichmann.com	jumbo.com	offerup.com	stockx.com	zalando-lounge.nl
backmarket.fr	depop.com	kelkoogroup.net	office.co.uk	studio.co.uk	zalando-privé.fr
banggood.com	dhgate.com	kiabi.com	orientaltrading.com	stylight.de	zalando-privé.it
barnesandnoble.com	dillards.com	kohls.com	otto.de	taobao.com	zalando.co.uk
baur.de	dirk.nl	kruidvat.nl	otto.nl	target.com	zalando.de
bax-shop.nl	dollartree.com	ladenzeile.de	overstock.com	thalia.de	zalando.fr
bazarchic.com	dsw.com	landsend.com	payback.de	the-saleroom.com	zalando.it
bcc.nl	e-leclerc.com	laredoute.fr	payline.com	therealreal.com	zalando.nl
bedbathandbeyond.com	ebay- kleinanzeigen.de	leenbakker.nl	picclick.com	thredup.com	zappos.com
belk.com	ebay.co.uk	letgo.com	picclick.de	tjx.com	zazzle.com
bergfreunde.de	ebay.com	libelle.nl	plus.nl	tkmaxx.com	zulily.com
bestsecret.com	ebay.de	libraccio.it	poco.de	topcashback.co.uk	getbitly.com
bever.nl	ebaydesc.com	lidl-flyer.com	poshmark.com	topshop.com	pickmypostcode.com
bidoo.com	emp.de	lidl-shop.nl	preis.de	tractorsupply.com	
biglots.com	etos.nl	lidl.de	privalia.com	trovit.it	

Tabella 11: Elenco siti Multibrand individuati

Siti Monobrand		
abercrombie.com	express.com	partycity.com
adidas.co.uk	forever21.com	patagonia.com
adidas.com	gap.com	peek-cloppenburg.de
adidas.it	halloweencostumes.com	prettylittlething.com
adidas.nl	hm.com	primark.com
ae.com	jcrew.com	rakuten.de
alight.com	jysk.nl	riverisland.com
anysearchmanager.com	kwantum.nl	romwe.com



**Politecnico
di Torino**

astegiudiziarie.it	lafeltrinelli.it	rosegal.com
bonprix.de	limango.de	shein.co.uk
bonprix.fr	lbean.com	shein.com
bonprix.it	loft.com	shop.lululemon.com
bonprix.nl	lululemon.com	shop.mango.com
boohoo.com	lulus.com	soliver.de
c-and-a.com	macys.com	spirithalloween.com
camaieu.fr	madewell.com	teespring.com
clarks.co.uk	mango.com	uniqlo.com
customink.com	missguided.co.uk	victoriassecret.com
dhlparcel.nl	mondadoristore.it	vsco.co
eddiebauer.com	newlook.com	yodel.co.uk
esprit.de	nike.com	zara.com

Tabella 12: Elenco siti Monobrand individuati



Bibliography

- Acemoglu, D., et al. 2019.** *Too much data: prices and inefficiencies in data market.* 2019.
- Acquisti A., Brandimarte L., Lowenstein G. 2015.** *Privacy and human behavior in the age of information.* 2015.
- Aridor, G., Che, Y. e Salz, T. 2020.** *The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR.* 2020.
- Batikas, M., et al. 2020.** *European Privacy Laws, Global Competition and the Privacy Landscape.* 2020.
- Bourreau, et al. 2020.** *Google/Fitbit will monetise health data and harm consumers.* 2020.
- Carvalho, et al.** *Protecting Citizens' Personal Data and privacy: Joint Effort from GDPR EU Cluster Reserach Projects.*
- Ding, X., et al. 2015.** *A fair data market system with data quality evaluation and repairing recommendation.* 2015.
- Fast, V., Schnurr, D. e Wohlfarth, M. 2018.** *The Value of Data in Digital Markets: Short-Term competitive Advantage or Long Term Market Power.* 2018.
- Fatehi, F., et al. 2020.** *General Data Protection Regulation (GDPR) in healthcare: Hot topics and reserch fronts.* 2020.
- Gibbons, E. M., et al. 2019.** *Antitrust Enforcement for the 21st Century.* 2019.
- Jiao, Y., et al. 2017.** *Profit maximization auction and data management in big data markets.* 2017.
- Kalan, R. S. 2020.** *Decentralization: a confluence of Data and Digital Personality.* 2020.
- Li, C., et al. 2013.** *A theory of pricing private data.* 2013.
- Morley, J., et al. 2022.** *Governing Data and Artificial Intelligence for Health Care: Developing an International Understanding.* 2022.
- Nicita A., Del Mastro M. 2020.** *Big Data .* 2020.
- Peng, M., et al. 2020.** *A cross-blockchain based distributed crowdsourcing system for internet of things.* 2020.
- Sabatino, L. e Sapi, G. 2021.** *Online Privacy and Market Structure: Theory and Evidence.* 2021.
- Sabatino, L., Congiu, R. e Sapi, G. 2022.** *The Impact of Privacy Regulation on Web Traffic: Evidence from the GDPR.* 2022.
- Song, M. 2022.** *Market Competition and regulatory compliance in public, non-profit, and for-profit organizations.* 2022.
- Subcommittee on Antitrust US. 2020.** *Investigation of Competition in Digital Markets.* 2020.
- Tatar, U., Gokce, Y. e Nussbaum. 2020.** *Laws versus technology: Blockchain, GDPR and tough tradeoffs.* 2020.
- Thai, M. T., Wu, W. e Xiong, H. 2016.** *Big Data in complex and social netoworks.* 2016.



**Politecnico
di Torino**

Verma, A. 2017. *Big data big deal*. 2017.



**Politecnico
di Torino**