



**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 degli effetti del GDPR sul traffico web,
tramite un confronto anagrafico e demografico del comportamento degli
utenti

Relatori:

Chiar.mo Prof. *Carlo Cambini*

Correlatore:

Chiar.mo Prof. *Lorien Sabatino*

Candidati:

Giovanni Gilardi

Mattero Fosforo



**Politecnico
di Torino**



**Politecnico
di Torino**

Ringraziamenti

I ringraziamenti vanno innanzitutto alla mia famiglia, che mi ha supportato durante tutto il mio percorso, e senza i quali sicuramente non sarei giunto a questo traguardo.

In secondo luogo, ai miei migliori amici Brondi, Caste, Forno, Edo, Mauri, Cino e Albi grazie ai quali ho trovato momenti di “pausa” dagli impegni universitari e con i quali la risata non manca mai.

Menzione speciale va agli amici conosciuti sui banchi universitari, compagni di progetti, giornate di studio, serate e vacanze: Leo, Peppe, Andre, Gloria, Fede, Chiara, Alessia, Erica e Daniel.

Infine, grazie a Giovanni, con cui ho condiviso non solo progetti, ma anche il percorso di stesura di questo documento e senza il quale sicuramente non sarei giunto fino a questo punto.



Politecnico
di Torino

Abstract

Il presente documento esamina gli effetti del General Data Protection Regulation (GDPR) sul traffico web, in particolare relativamente ai siti appartenenti alle categorie “News&Media” e “Social Network”. L’analisi è svolta tramite un confronto a livello anagrafico, suddividendo gli utenti in sei diverse fasce d’età, e demografico, per genere. Per ogni dominio viene individuata la fascia d’età di provenienza della maggior parte degli utenti e il genere prevalente. Si confronta il loro comportamento prima e dopo l’entrata in vigore del GDPR, svolgendo una comparazione tra il mondo europeo e statunitense. I risultati ottenuti tramite la prima analisi svolta con Microsoft Excel dimostrano come le donne abbiano una propensione maggiore a essere influenzate da tematiche relative alla privacy, pur riscontrando una riduzione media nel traffico per entrambi i generi, così come le persone di età compresa tra 25 e 64 anni. Un secondo risultato indica come sia modificata soltanto la prima iterazione tra dominio e utente, rilevando una diminuzione di traffico medio e un aumento della frequenza di rimbalzo, mentre durata delle visite e numero medio di pagine visitate non subiscono effetti significativi.

L’implementazione dell’analisi sul Software stata al contrario riporta come l’unica variazione statisticamente significativa sia relativa alla durata media delle visite degli uomini che risulta ridotta.



Sommario

Indice figure	7
Indice tabelle	8
1. Introduzione: dati digitali e privacy online nell'era della connessione	10
2. Literature Review	13
2.1 Approccio metodologico	13
2.2 Literature review: data market	15
2.2 Literature review: GDPR	21
2.2.1 Introduzione al GDPR	21
2.2.2 Approccio metodologico per la literature review	22
2.3 Literature review: Governance and Regulation	27
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	46
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	50
4.3 Statistiche descrittive	52
4.4 Statistiche temporali	61
4.4.1 Statistiche di confronto: Media e Social	65
4.4.2 Statistiche relative al confronto per genere	68



**Politecnico
di Torino**

4.4.3	Statistiche relative al confronto anagrafico	73
5.	<i>Modelli econometrici</i>	82
6.	<i>Conclusioni e riflessioni finali</i>	87
Appendice		90
Appendice 1		90
Appendice 2		93
Bibliography		103



Indice figure

Figura 1: approccio metodologico della literature review.....	14
Figura 2: approccio metodologico alla literature review sul mercato dei dati digitali...	15
Figura 3: Data Market Co-occurrence Map	16
Figura 4: Mappa riassuntiva delle principali iniziative del GDPR (Carvalho, et al.)	22
Figura 5: approccio metodologico alla literature review sul GDPR.....	23
Figura 6: GDPR Co-occurrence Map.....	24
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



Indice tabelle

Tabella 1: Market Share, Devices & App Stores	18
Tabella 2: Market Share, Search Industry	19
Tabella 3: Market Share, E-commerce	19
Tabella 4: Market Share, Social Media	20
Tabella 5: suddivisione Media e Social	52
Tabella 6: suddivisione per nazione di provenienza del traffico	53
Tabella 7: suddivisione per nazione di provenienza del traffico diviso per categorie ...	54
Tabella 8: dimensione dei siti News&Media	55
Tabella 9: dimensione dei siti Social Network.....	55
Tabella 10: suddivisione per genere	56
Tabella 11: suddivisione per fascia d'età	57
Tabella 12: suddivisione per seconda fascia d'età	57
Tabella 13: suddivisioni per genere Social e Media.....	58
Tabella 14: suddivisioni per fasce d'età Social e Media.....	59
Tabella 15: suddivisioni per genere tra Europa e USA.....	60
Tabella 16: suddivisioni anagrafiche tra Europa e USA	61
Tabella 17: confronto traffico medio tra Europa e USA	62
Tabella 18: confronto durata visite tra Europa e USA	63
Tabella 19: confronto pagine visitate tra Europa e USA.....	63
Tabella 20: confronto Bounce Rate tra Europa e USA.....	64
Tabella 21: Confronto traffico medio Media tra Europa e USA	65
Tabella 22: Confronto bounce rate medio Media tra Europa e USA.....	66
Tabella 23: confronto traffico medio Social tra Europa e USA	66
Tabella 24: confronto bounce rate medio Social tra Europa e USA	67
Tabella 25: confronto traffico medio tra uomini e donne europei	68
Tabella 26: confronto bounce rate tra uomini e donne europei	69
Tabella 27: confronto traffico medio degli uomini tra Europa e USA	70
Tabella 28: confronto Bounce Rate degli uomini tra Europa e USA.....	70
Tabella 29: confronto traffico medio delle donne tra Europa e USA	71
Tabella 30: confronto Bounce Rate delle donne tra Europa e USA	72
Tabella 31: confronto traffico tra le fasce d'età a livello europeo	73
Tabella 32: confronto durata visite tra fasce d'età a livello europeo.....	74
Tabella 33: confronto Bounce Rate tra le fasce d'età a livello europeo	74
Tabella 34: confronto traffico medio 18-24 anni tra Europa e USA	75



Tabella 35: confronto Bounce Rate 18-24 anni tra Europa e USA	76
Tabella 36: confronto traffico medio 25-34 anni tra Europa e USA	77
Tabella 37: confronto Bounce Rate 25-34 anni tra Europa e USA	77
Tabella 38: confronto traffico medio 35-44 anni tra Europa e USA	78
Tabella 39: confronto Bounce Rate 35-44 anni tra Europa e USA	79
Tabella 40: confronto traffico medio 65+ anni tra Europa e USA	80
Tabella 41: confronto Bounce Rate 65+ anni tra Europa e USA.....	80
Tabella 42: regressione lineare relativa al traffico	83
Tabella 43: regressione lineare relativa al tempo di visita	84
Tabella 44: regressione lineare relativa al numero di pagine visitate	85
Tabella 45: regressione lineare relativa al bounce rate.....	86
Tabella 46: elenco siti News&Media	92
Tabella 47: elenco siti Social Network	92
Tabella 48: Confronto durata visite News&Media tra Europa e USA.....	93
Tabella 49: Confronto pagine visitate News&Media tra Europa e USA	93
Tabella 50: Confronto durata visite Social Network tra Europa e USA	94
Tabella 51: Confronto pagine visitate Social Network tra Europa e USA	94
Tabella 52: Confronto differenziale durata visite tra uomini e donne europei	95
Tabella 53: Confronto differenziale pagine visitate tra uomini e donne europei.....	95
Tabella 54: Confronto durata visite uomini tra Europa e USA.....	96
Tabella 55: Confronto pagine visitate uomini tra Europa e USA.....	96
Tabella 56: Confronto durata visite donne tra Europa e USA.....	97
Tabella 57: Confronto pagine visitate donne tra Europa e USA.....	97
Tabella 58: confronto pagine visitate tra fasce d'età a livello europeo	98
Tabella 59: confronto durata visite 18-24 anni tra Europa e USA.....	98
Tabella 60: confronto durata visite 18-24 anni tra Europa e USA.....	99
Tabella 61: confronto durata visite 25-34 anni tra Europa e USA.....	99
Tabella 62: confronto pagine visitate 25-34 anni tra Europa e USA	100
Tabella 63: confronto durata visite 35-44 anni tra Europa e USA.....	100
Tabella 64: confronto pagine visitate 35-44 anni tra Europa e USA	101
Tabella 65: confronto durata visite 65+ anni tra Europa e USA.....	101
Tabella 66: confronto pagine visitate 65+ anni tra Europa e USA.....	102



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

Come si può evincere dunque dalla Fig. 2 è stata fatta una ricerca per parole chiave sulla base delle seguenti keywords: TITLE-ABS-KEY ("market" AND "data" AND "competition" AND "privacy").

Questo ha permesso di individuare un cluster di 106 documenti¹ 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 3 individuando 79 parole chiave riportate all'interno della co-occurrence map in Fig. 3.

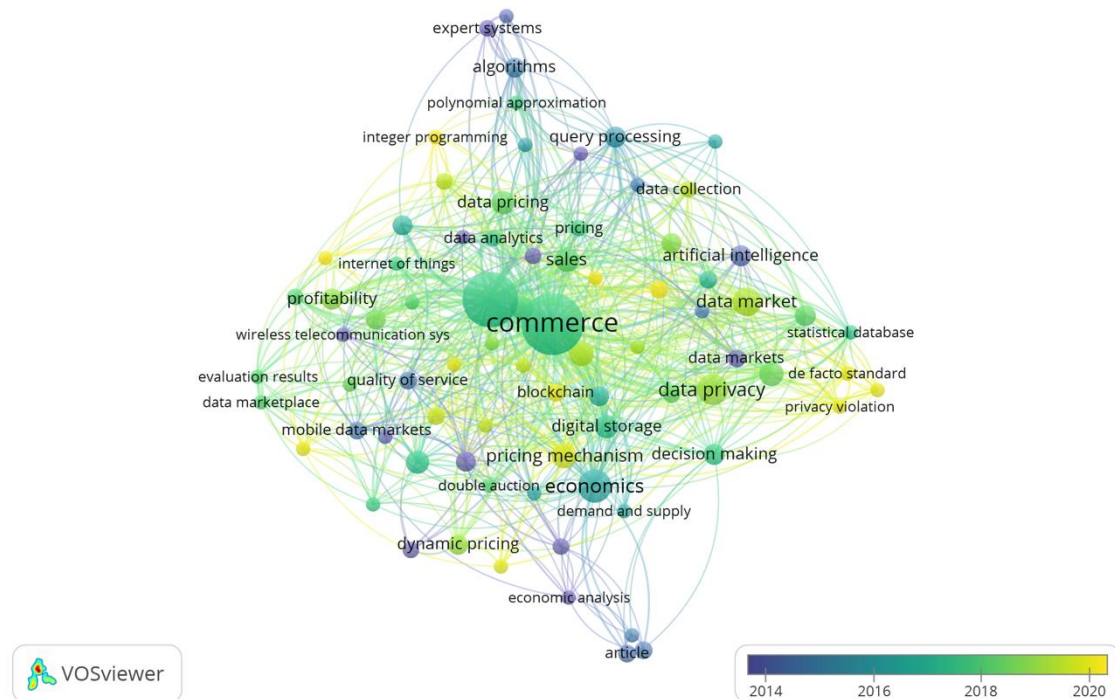


Figura 3: Data Market Co-occurrence Map

Come si può vedere dalla network map riportata in Fig. 3 la ricerca scientifica negli ultimi anni ha toccato diverse tematiche e argomenti ed è particolarmente semplice individuare il trend che ha percorso in funzione anche dell'evoluzione del mercato stesso.

¹ si riporta il [link](#) al documento estratto da SCOPUS con i 106 documenti analizzati per creare la co-ccurrence relativa al "Data market"



Si inizi con il fornire alcune chiavi di lettura per rendere più semplice la visione e la comprensione della mappa stessa:

- 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.

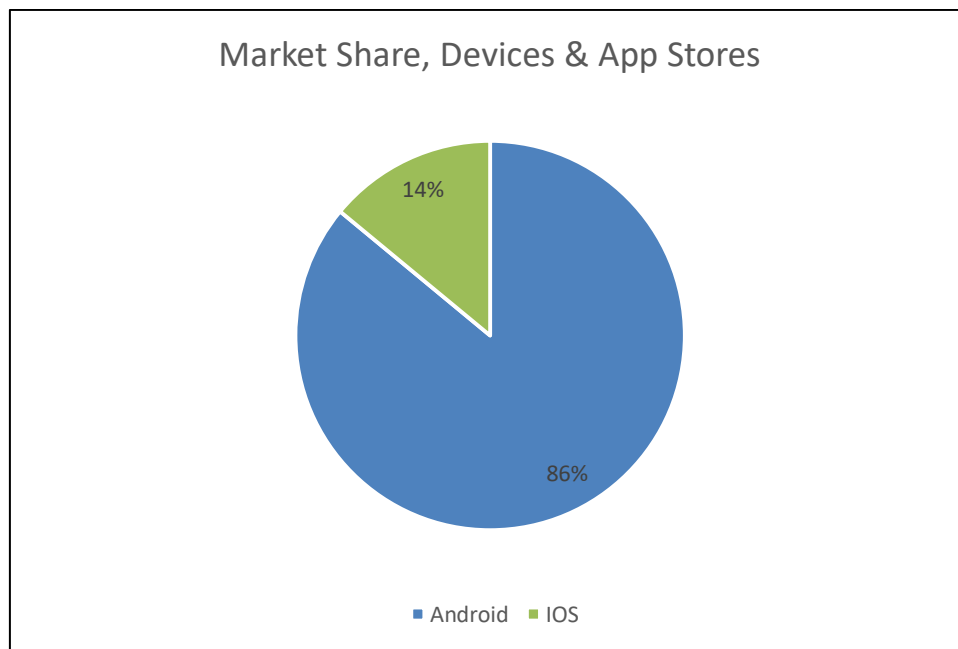


Tabella 1: Market Share, Devices & App Stores



Market Share, Search Industry

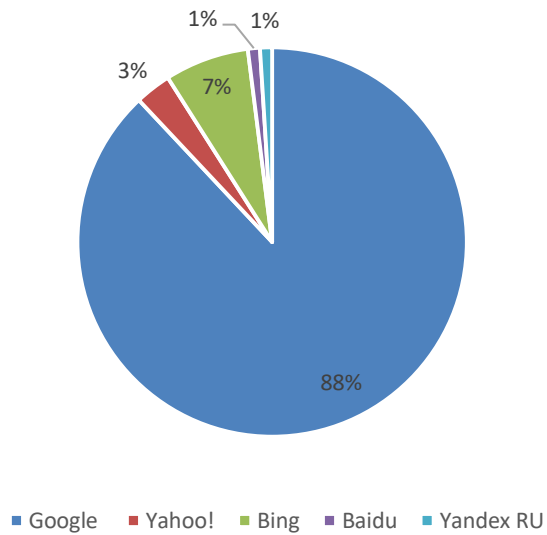


Tabella 2: Market Share, Search Industry

Market Share, E-commerce

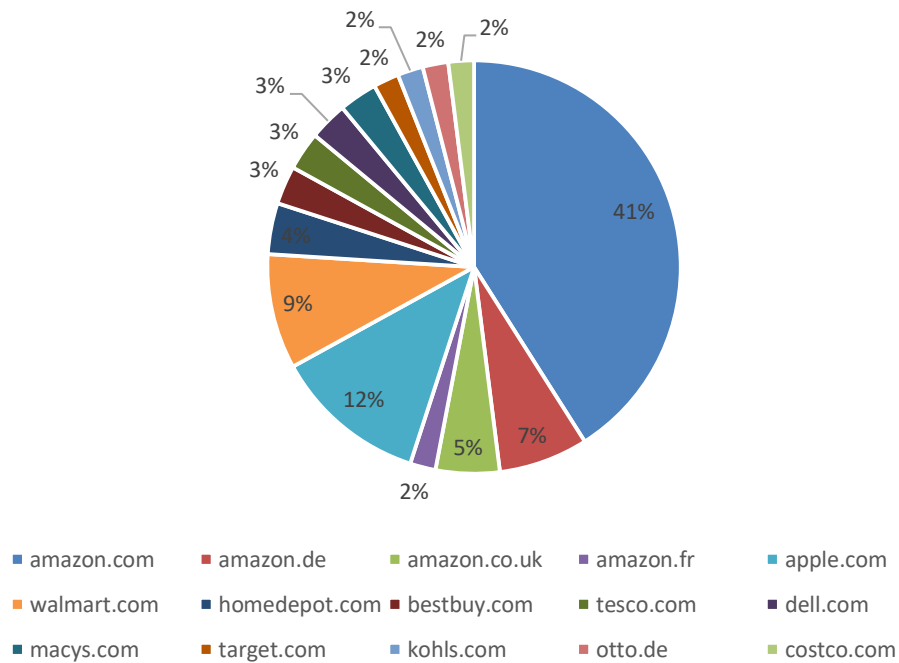


Tabella 3: Market Share, E-commerce

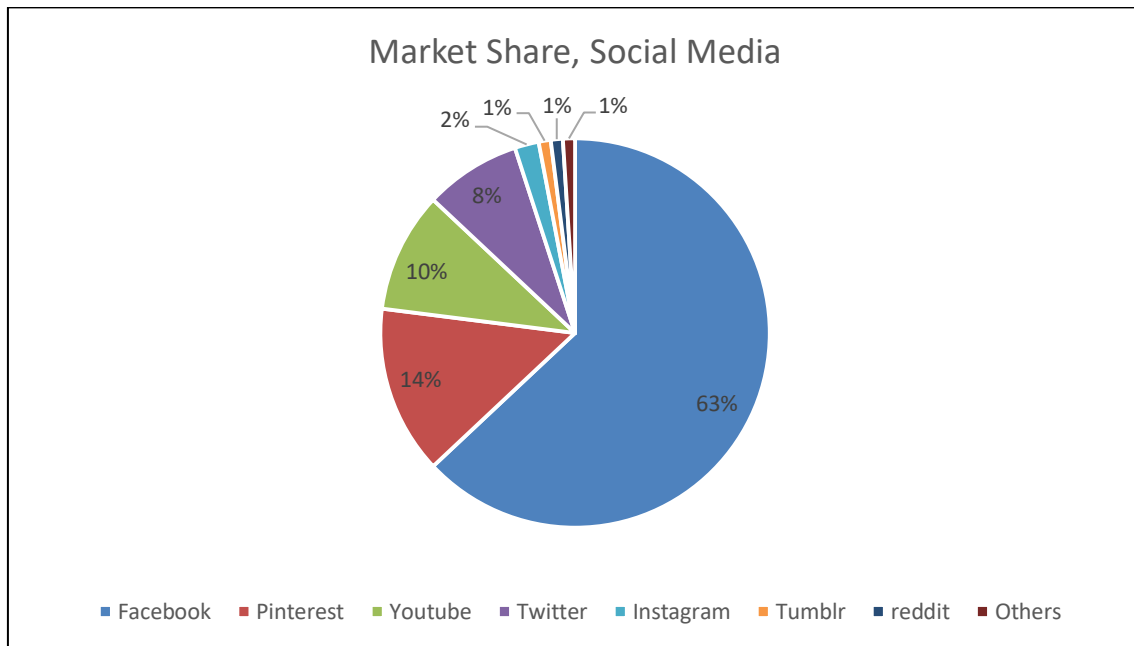


Tabella 4: Market Share, Social Media

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 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.



2.2 Literature review: GDPR

2.2.1 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

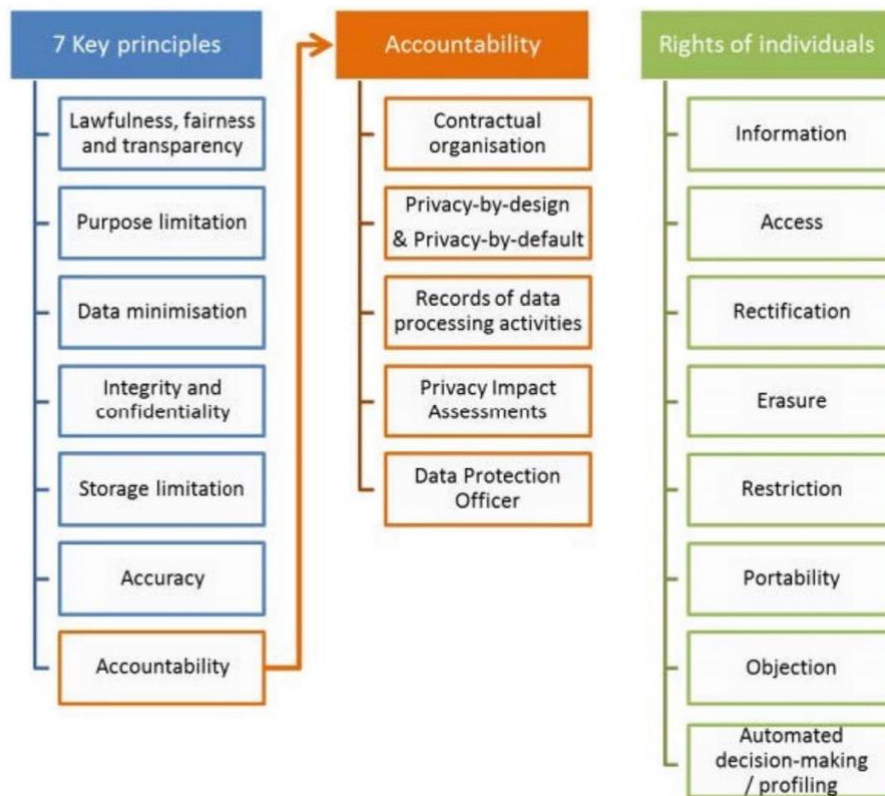


Figura 4: Mappa riassuntiva delle principali iniziative del GDPR (Carvalho, et al.)

2.2.2 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.

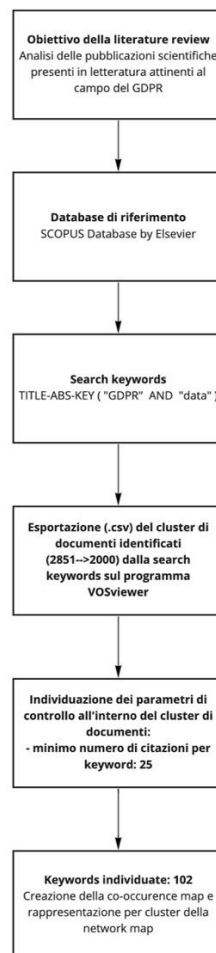


Figura 5: approccio metodologico alla literature review sul GDPR

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.

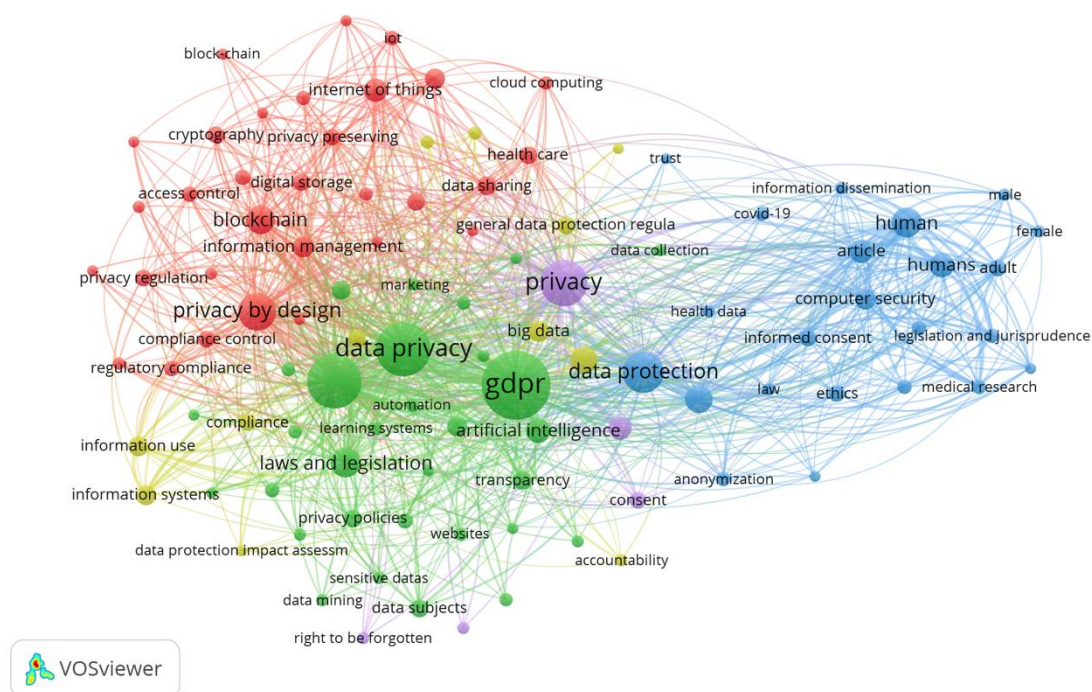


Figura 6: GDPR Co-occurrence Map

Analizzando la mappa riportata in Fig. 5, è facile individuare i 5 insiemi che il software VOSviewer ha identificato (3 più rilevanti e 2 meno rilevanti) dallo studio dei 2000 documenti selezionati:

1. In rosso vengono riportate tutte le tematiche affrontate dalla letteratura scientifica inerenti alle tecnologie innovative che oggi il Garante sta analizzando

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



- con occhio estremamente attento come: blockchain, IoT, cloud computing e cryptography
2. In verde vengono riportati i temi trattati all'interno delle pubblicazioni scientifiche relative al contenuto del GDPR e quindi agli obiettivi che si è preposto il garante introducendo questo regolamento (es. data privacy, laws and legislation, websites, transparency)
 3. In blu vengono riportati i campi su cui il GDPR ha avuto maggiore impatto relativamente al singolo individuo: protezione dei dati personali (data protection), maggiore fiducia ed etica nella condivisione ed utilizzo dei dati personali
 4. In giallo vengono riportati alcuni dei principi introdotti dal GDPR, come quello di accountability per il responsabile del trattamento (trattato all'interno del Cap. 2.2.1)
 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.

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.

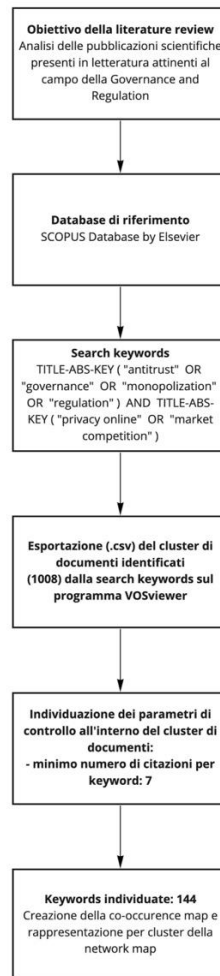


Figura 7: approccio metodologico alla literature review sulla Governance and Regulation

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-occurrence map in Fig. 8.

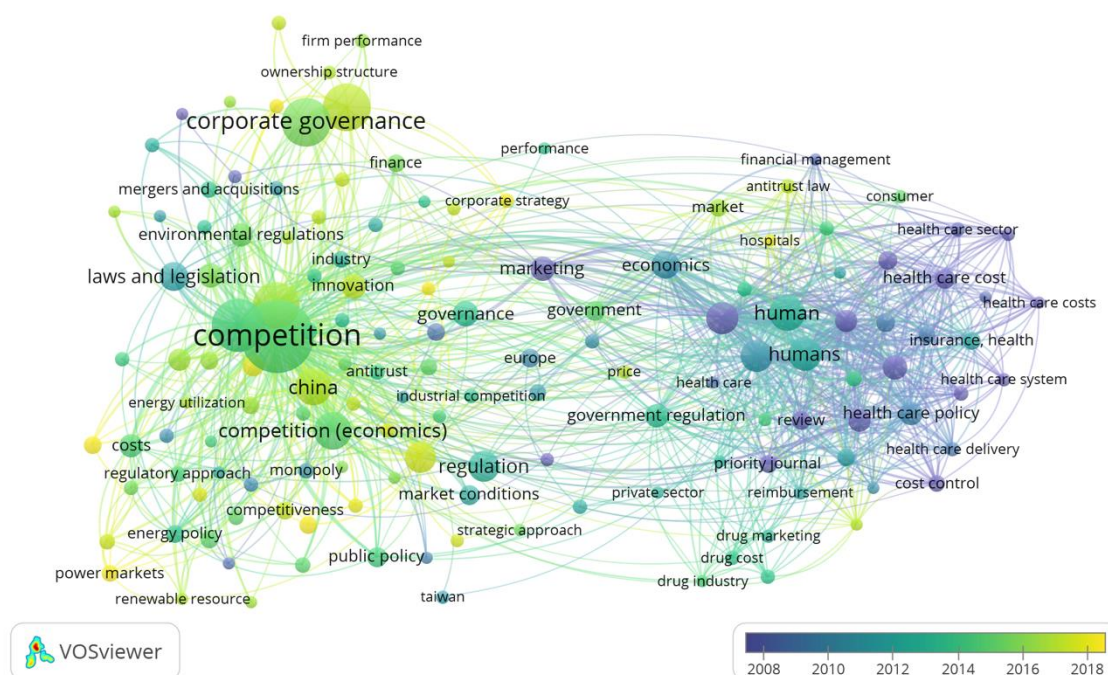


Figura 8: Co-occurrence Map Governance and Regulation

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

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



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.

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



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).

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

TITLE-ABS-KEY ("privacy" AND "online" AND "big data" AND "digital").

Una volta caricato il file su VOSviewer si è impostata una threshold relativa al numero minimo di citazioni per le keywords pari a 5 individuando 54 parole chiave riportate all'interno della co-occurrence map in Fig. 10.

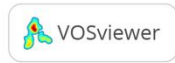


Figura 10: Co-occurrence Map Privacy

Come si può vedere dalla co-occurrence map riportata in Fig. 10 la ricerca scientifica negli anni passati ha prodotto pubblicazioni ed articoli in un orizzonte temporale più ristretto rispetto ai casi precedenti in quanto l'estrazione da SCOPUS ha individuato documenti pubblicati tra il 2012 e il 2022.

5 si riporta il [link](#) al documento estratto da SCOPUS con i 1008 documenti analizzati per creare la co-ccurrence map relativa alla “Privacy”



All'interno di questo intervallo però le keyword che hanno rispettato la threshold introdotta fanno riferimento a documenti compresi tra il 2017 e il 2019.

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 capitolo 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 al mondo di News & Media e Social Network.

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 proveniente 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.

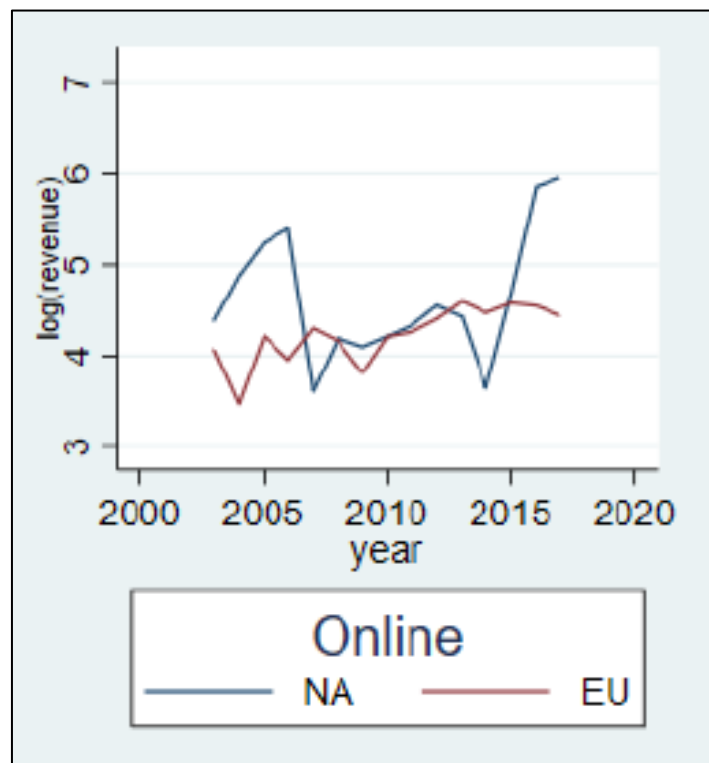


Figura 11: Trend di variazione revenues online market



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.

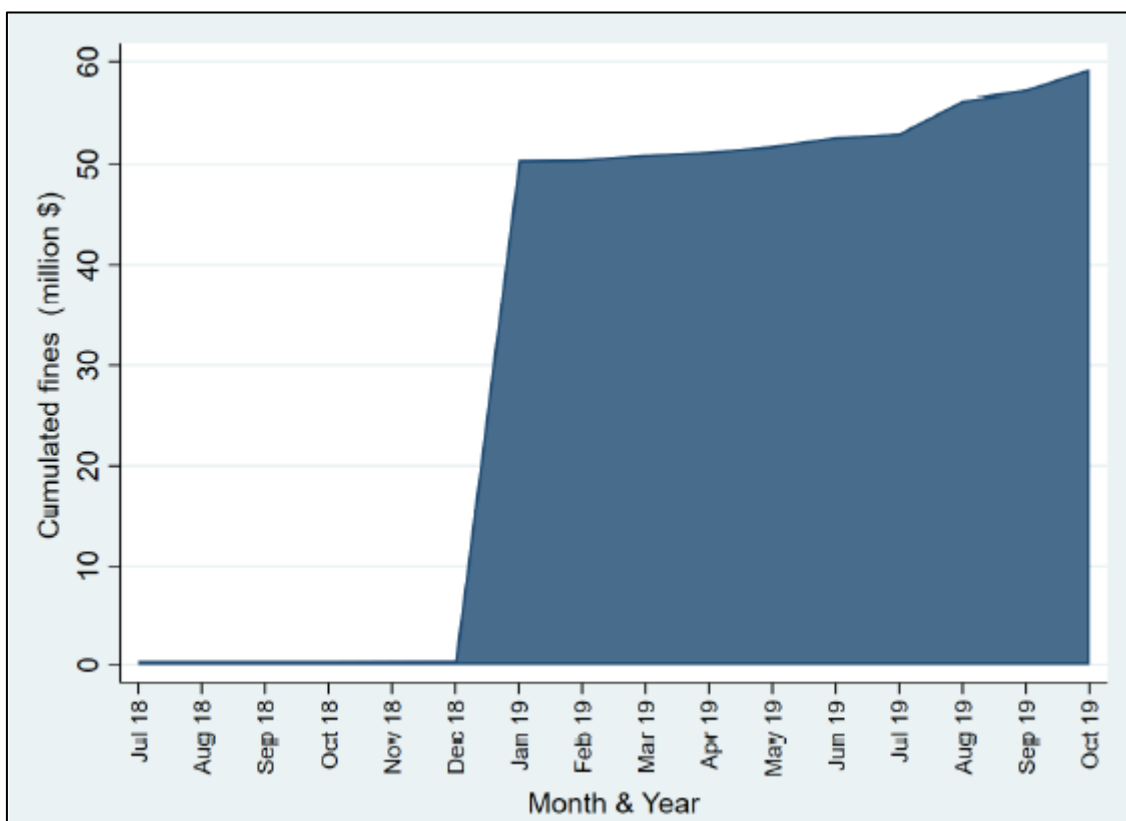


Figura 12: Cumulata delle sanzioni dovute al GDPR



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 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’industry 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.

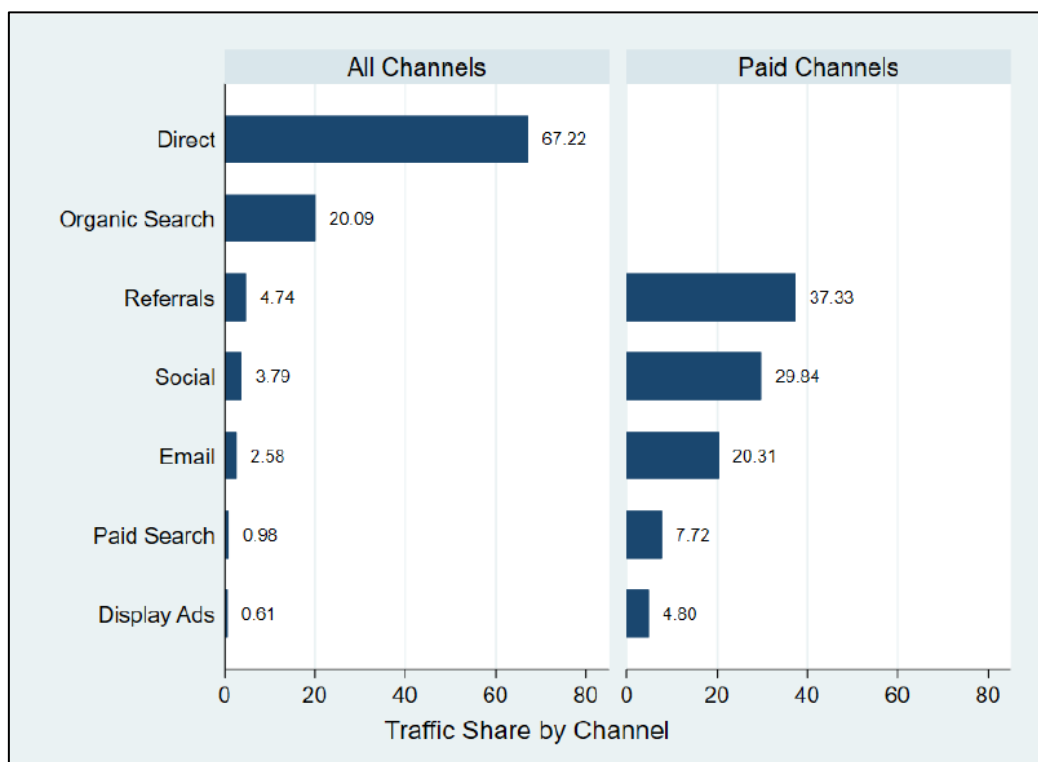


Figura 13: Distribuzione del traffico online per canale

Un possibile problema nei dati poteva riguardare il fatto che a fronte delle norme presenti nel regolamento, il provider di dati 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 provider di dati 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.

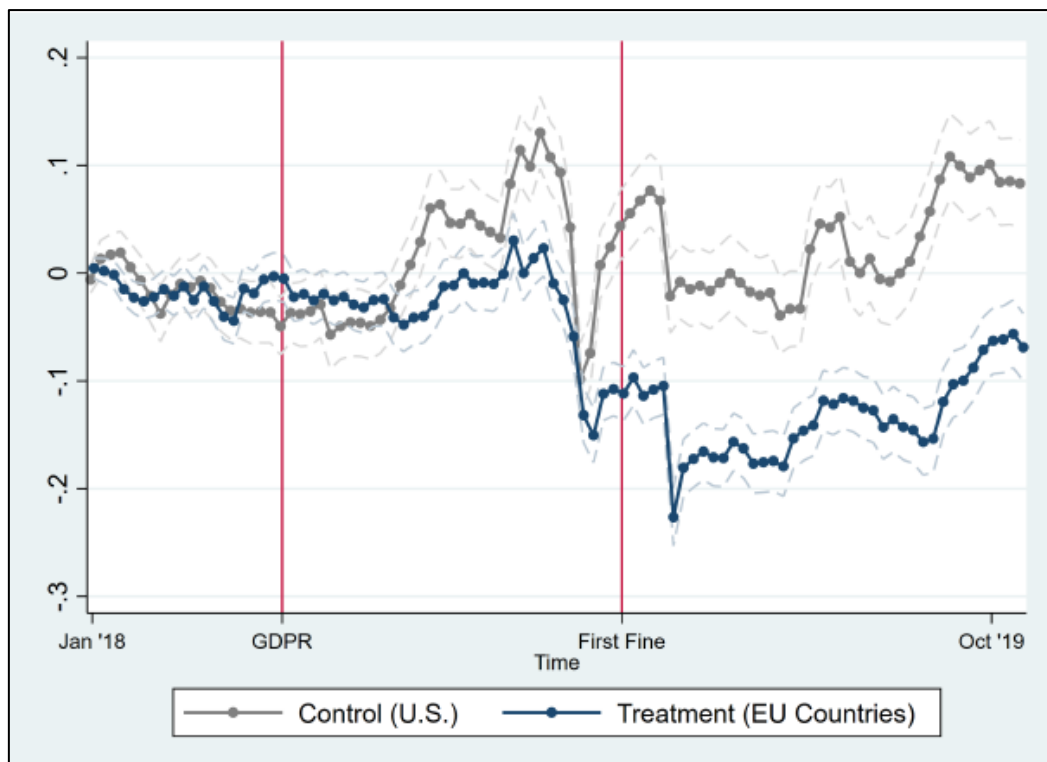


Figura 14: Variazione del traffico web

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.

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

Partendo dall'analisi della letteratura precedente relativa agli effetti della direttiva ePrivacy del 2009 e del regolamento europeo, il GDPR, si è deciso di procedere a una ricerca specifica in alcuni ambiti di analisi circoscritti.

In particolare, l'obiettivo è valutare la differenziazione tra categorie di utenti su due basi:

- **Demografica**, cioè valutare come il regolamento introdotto ha avuto effetti diversi tra uomini e donne;
- **Anagrafica**, studiando se e come i comportamenti si siano modificati eterogeneamente tra le diverse fasce d'età degli utenti, sempre a seguito dell'entrata in vigore del GDPR.

Per fare ciò, all'interno della totalità dei siti presenti sono stati selezionati quelli appartenenti alla categoria "News&Media" e alla categoria "Social Network", categorie che al giorno d'oggi possono essere considerate le fonti principali tramite cui le informazioni raggiungono le persone.

La scelta di questa restrizione deriva dal fatto che le tipologie di siti selezionate raggiungono la quasi totalità degli utenti, in quanto l'informazione non rappresenta un argomento ristretto ad alcune tipologie di persone. In tal modo, i risultati non saranno influenzati da tutti quei domini o quelle categorie che al contrario hanno come target fasce specifiche di utenti, e avrebbero quindi potuto creare deviazioni.

La scelta degli approfondimenti, permetterà di capire come le diverse categorie di persone abbiano reagito di fronte all'introduzione del nuovo regolamento sulla privacy, in quanto si potrà sia svolgere un confronto diretto tra le categorie, ovviamente focalizzandosi a livello europeo in questa prima analisi, ma anche andare a confrontare i comportamenti di soggetti appartenenti alla stessa categoria, ma residenti in parti del mondo diverse, ottenendo un confronto puntuale tra coloro soggetti alla nuova normativa, cioè i cittadini europei, e coloro che invece non lo sono.



4.2 Introduzione al dataset

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.

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. E-mail

Il primo passo è stato ridurre il dataset, considerando soltanto i siti appartenenti alla categoria “News_and_Media” oppure “Social_Networks_and_Online_Communities” in modo tale da lavorare soltanto su tali domini.

A questo punto, grazie ai dati prodotti da Un tool di analisi online, sono state inserite tre colonne aggiuntive:

- **“Fascia di età con più visite”**: l’intervallo anagrafico in cui ricadeva la maggior parte degli utenti del sito



- **“Seconda fascia d’età”**: l’intervallo anagrafico al secondo posto per numero di utenti
- **“Uomini/Donne”**: una variabile dummy presentante valore 0 se la maggioranza (>60%) degli utenti del sito sono donne, valore 2 se la maggioranza (>60%) sono uomini, o valore 1 qualora non si presentasse una maggioranza netta e quindi si potesse considerare all’incirca pari l’affluenza tra uomini e donne

A valle di queste modifiche il dataset presenta circa 60400 dati per ciascuna colonna, per un totale di 1,1M di dati singoli da utilizzare nell’analisi.

4.3 Statistiche descrittive

Nel database utilizzato per gli studi sono presenti una totalità di 637 coppie dominio-paese, questo in quanto per alcuni siti il traffico proviene da diversi paesi ed è quindi corretto non considerare soltanto i diversi domini come base per lo studio ma considerare la coppia in aggregato.

Tra questi possiamo effettuare una prima distinzione tra siti appartenenti al mondo dei Media, 505, oppure a quello dei Social, 132:

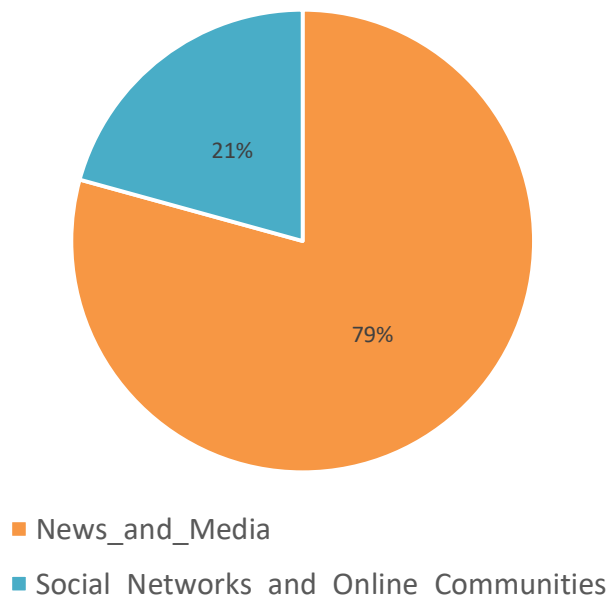


Tabella 5: suddivisione Media e Social



Come vediamo la maggioranza appartengono al mondo dei media, questo è spiegabile dalla diversa tipologia di pubblico e utenti raggiunti. Infatti, molti media hanno una risonanza prettamente nazionale, mentre i social raggiungono utenti da tutto il mondo e per questo il loro numero risulta minore. Un'altra possibile motivazione riguarda la diffusione temporale delle due diverse categorie, come tutti sappiamo i social network sono un fenomeno relativamente recente al contrario dei canali d'informazione.

Una seconda informazione rilevante deriva dal paese di provenienza del traffico:

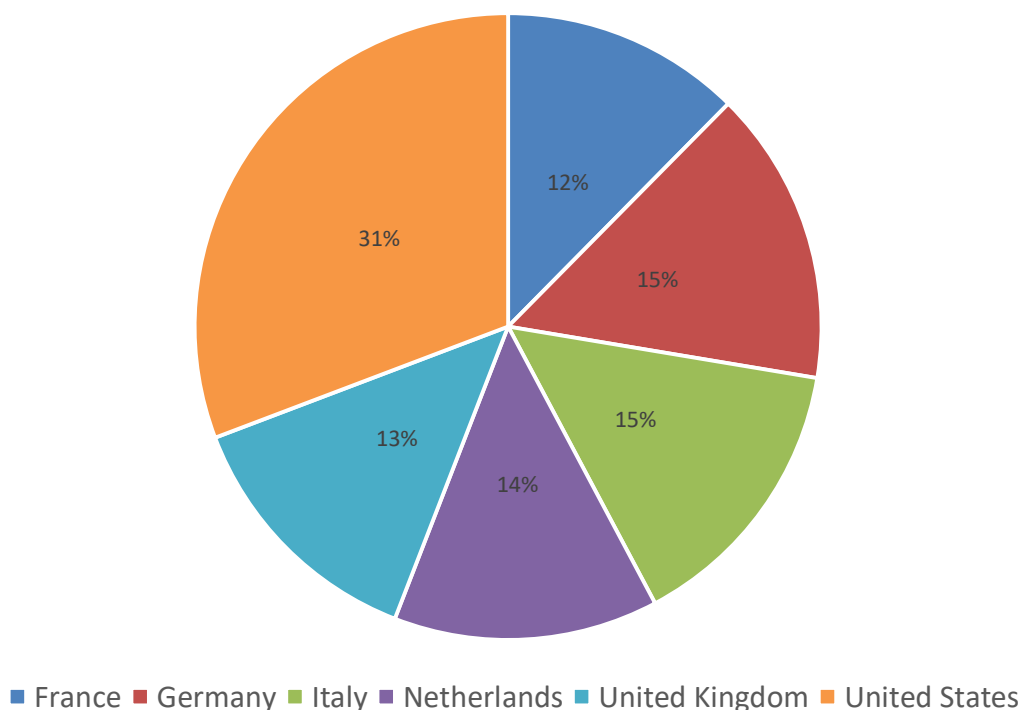


Tabella 6: suddivisione per nazione di provenienza del traffico

Come visibile la maggior parte, circa un terzo proviene dagli Stati Uniti, ma dato che i nostri studi andranno ad aggregare il traffico europeo in un unico dato possiamo concludere che la sproporzione visibile sia a favore dell'Europa, con circa i 2/3 del totale, questi valori sono congrui al tipo di analisi che verrà svolta, si effettueranno infatti due tipologie di confronto, la prima tra i dati europei delle diverse categorie che verranno comparati per valutare eventuali differenze significative, la seconda tra siti europei e statunitensi in modo da verificare che le differenze rilevate siano realmente imputabili al GDPR.



Il dato riguardante le provenienza geografica può essere analizzato singolarmente per le due categorie di domini presenti nel dataset.

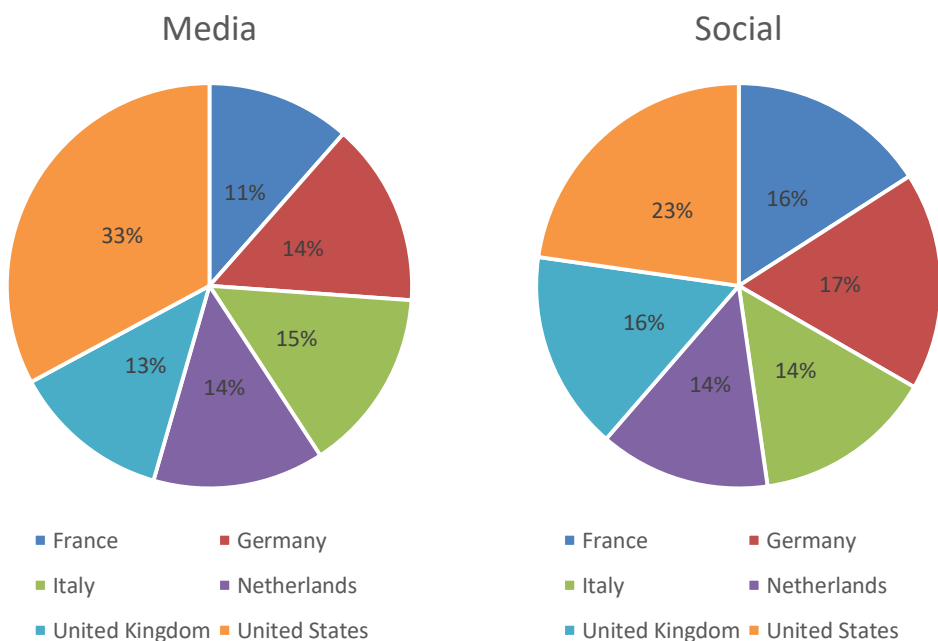


Tabella 7: suddivisione per nazione di provenienza del traffico diviso per categorie

Da quest'ultimo confronto emerge un dato molto interessante: la distribuzione geografica dei Social Network risulta essere molto equilibrata, mantenendo un valore leggermente più alto per gli Stati Uniti, ma in maniera nettamente inferiore rispetto ai Media. Questo porta a invertire il pensiero comune che colloca tutti i social come proveniente dallo stato americano.

Queste prime due statistiche permettono di fornire una panoramica generale della suddivisione dei dati presenti e delle loro fonti.

Per quanto riguarda la dimensione dei domini presenti, rilevata grazie al numero totale di visite ricevute da essi vengono riportati i seguenti grafici:



Dimensione siti News&Media

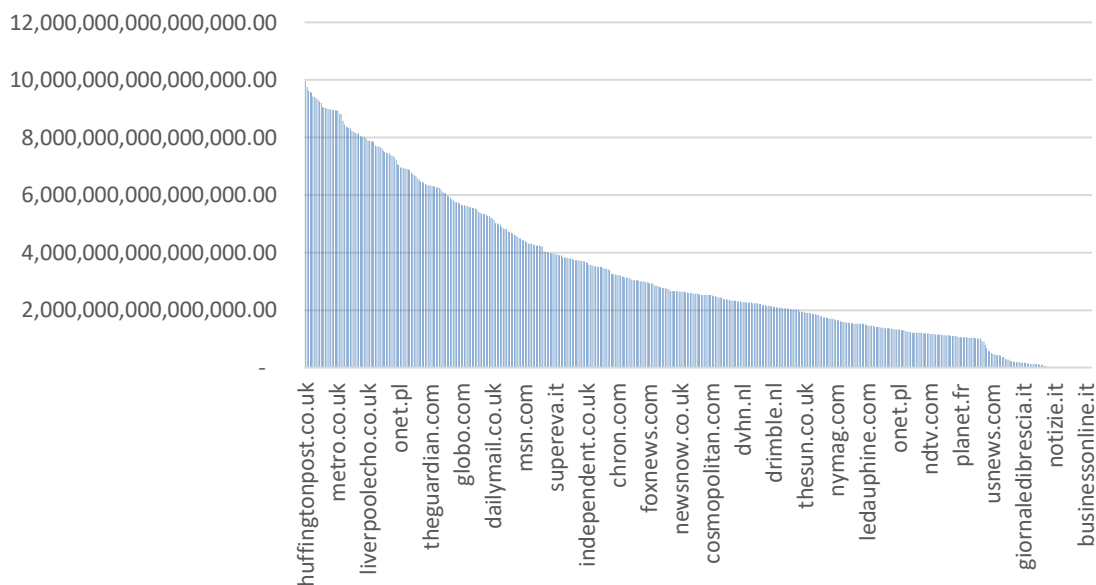


Tabella 8: dimensione dei siti News&Media

Dimensione siti Social Network

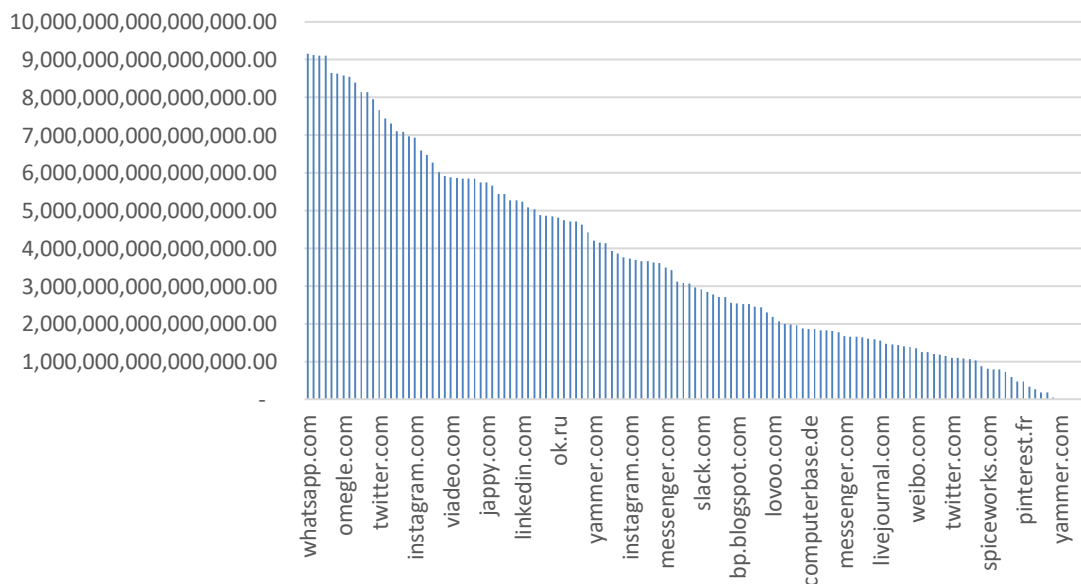


Tabella 9: dimensione dei siti Social Network



Si nota un'elevata eterogeneità nelle dimensioni, rendendo le analisi successive indipendenti dalla grandezza del dominio in esame, in quanto il campione risulta rappresentativo di tutte le tipologie.

Inoltre, si vede come contrariamente a ciò che si potrebbe pensare non risulta presentarsi uno squilibrio della dimensione dei siti a favore dei social, nemmeno per i più utilizzati come Whatsapp oppure Facebook, che anzi ottengono valori di visite comparabili ai più importanti siti d'informazione. È bene però specificare come i dati risultino soltanto relativi agli accessi tramite desktop, e non alle applicazioni mobili, questo condiziona fortemente alcuni social network, è primo tra tutti Whatsapp.

Per quanto riguarda gli elementi di studio, si riportano di seguito le statistiche relative a:

- Suddivisione dei siti sulla base del sesso della maggioranza degli utenti:

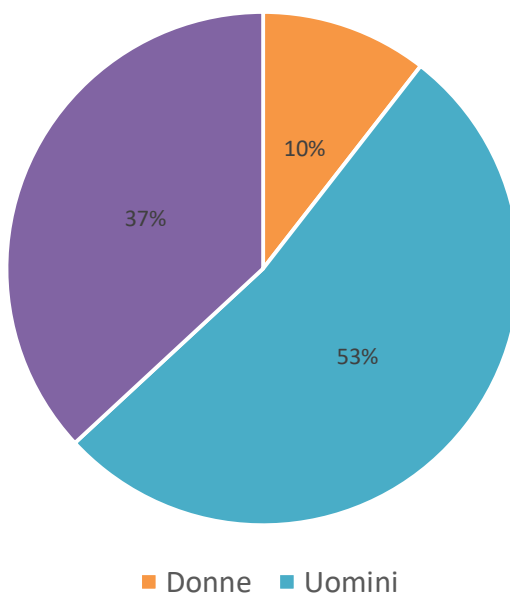


Tabella 10: suddivisione per genere

- Suddivisione anagrafica relativa alla prima fascia:



Politecnico
di Torino

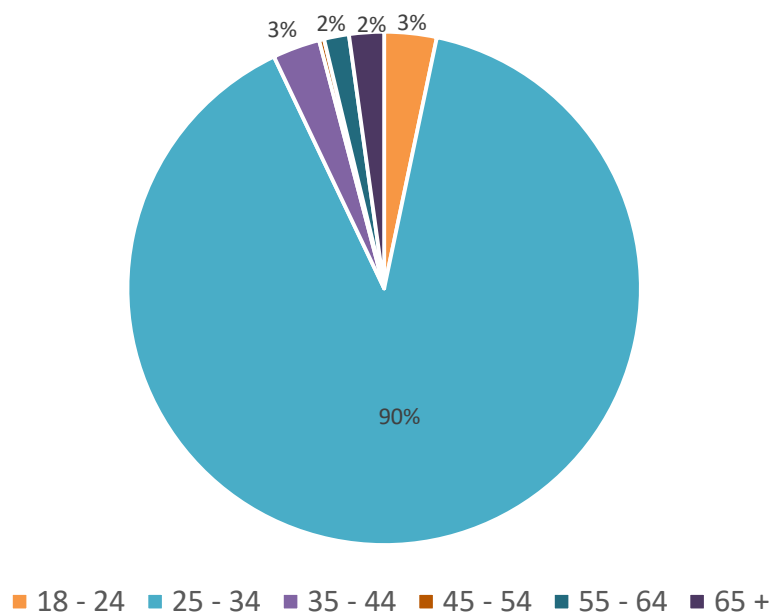


Tabella 11: suddivisione per fascia d'età

- Suddivisione anagrafica relativa alla seconda fascia:

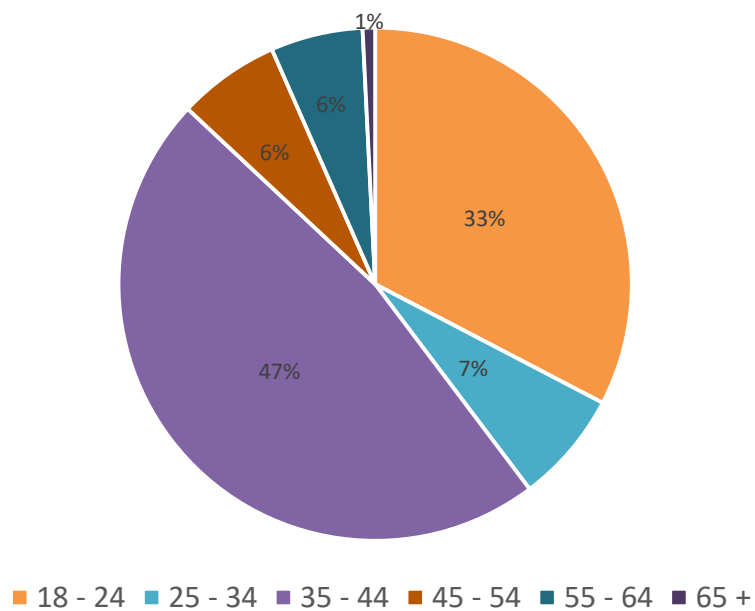


Tabella 12: suddivisione per seconda fascia d'età



Dai tre grafici sopra si evidenziano i primi risultati interessanti dello studio.

Partendo dalla suddivisione demografica notiamo come più del 50% della totalità dei siti presenti una prevalenza di uomini tra i proprio utenti, il 37% circa presenti un rapporto di parità, mentre solo il restante 13% veda prevalere le donne. La spiegazione di questa disuguaglianza può essere ritrovata nel contenuto presentato, infatti, i contenuti prettamente femminili risultano essere nettamente inferiori rispetto ai contenuti maschili. Più sorprendente risulta, invece, la maggioranza dei siti dove prevalgono gli uomini rispetto a quelli in cui il rapporto risulta bilanciato, in quanto all'interno di categorie di siti il cui scopo è l'informazione, essa dovrebbe presentare argomenti che portino interesse di tutta la popolazione, non soltanto della fascia maschile.

Queste proporzioni vengono mantenute anche andando a studiare singolarmente le due categorie in esame, riguardanti i social e i media come si può vedere dai grafici seguenti:

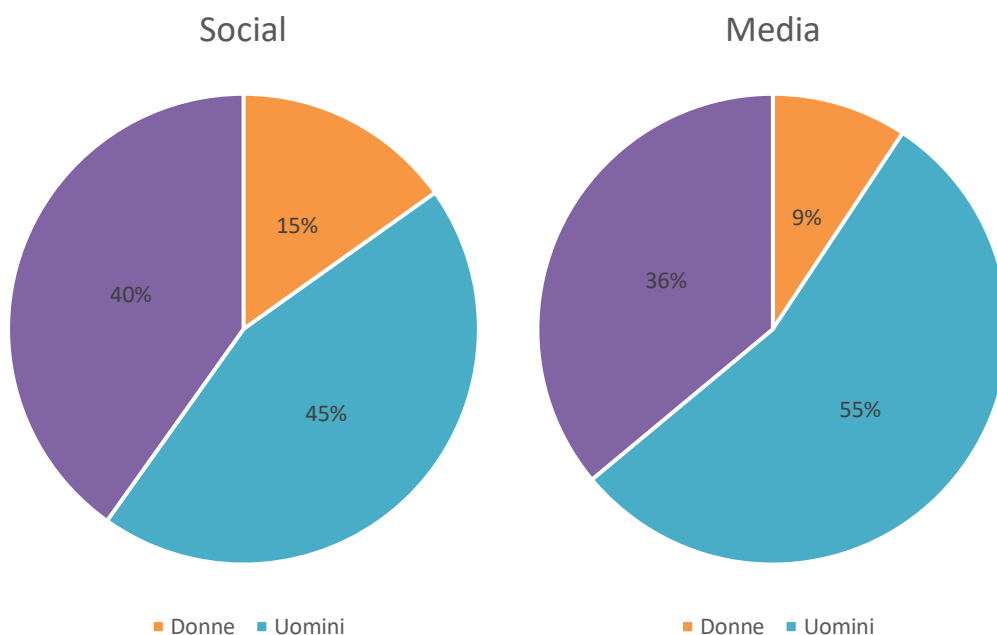


Tabella 13: suddivisioni per genere Social e Media

La disparità risulta essere ridotta, anche se non di molto, all'interno della categoria dei social, questo in quanto per definizione questa tipologia di siti trova il suo scopo nel mettere in contatto tra loro gli utenti, per fare ciò è necessario che non vi sia una disparità di genere troppo alta. Al contrario le statistiche sembrerebbero evidenziare come gli uomini siano più portati a recarsi su siti d'informazione rispetto alle donne.



Per quanto riguarda i dati demografici, il risultato più evidente è là che il 90% dei domini totali vede la percentuale più elevata di utenti all'interno della fascia compresa tra i 25 e i 34 anni, questo risultato è molto significativo per gli studi successivi in quanto consente di suddividere in due categorie anagrafiche i dati ottenuti ed applicarli così al modello di regressione. Questa evidente disparità può trovare la sua causa nel fatto che gli utenti appartenenti alla fascia prevalente sono coloro che hanno vissuto più da vicino l'avvento del mondo online durante la loro infanzia e adolescenza, e per questo sono coloro che tendono ad utilizzare prevalentemente questo strumento come fonte principale di informazioni. Questa teoria trova conferma anche nella suddivisione della seconda fascia di visite dove, le due categorie preponderanti, che unite raggiungono l'80% del totale, risultano essere la fascia tra i 18 e i 24 anni e quella tra 35 e 44, che sono i due intervalli anagrafici più influenzati dall'avvento di internet oltre a quello citato precedentemente. Sorprendente risulta, però, la bassa percentuale, nella prima fascia, di utenti tra 18 e 24 anni, una generazione che fa dell'uso di internet e dei social uno dei suoi punti cardine. La spiegazione a questo basso valore si può ipotizzare nella scarsa tendenza all'informazione di questa tipologia di utenti, e anche al fatto che i contenuti di molti Media sono rivolti a target di età maggiore.

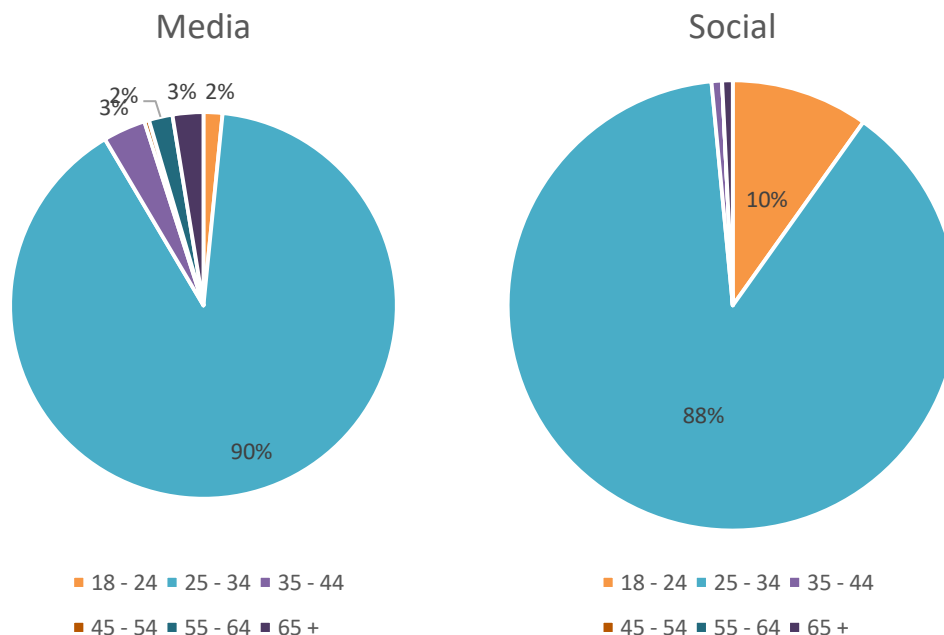


Tabella 14: suddivisioni per fasce d'età Social e Media

Le statistiche suddivise per tipologia di dominio confermano la teoria esposta in quanto vediamo come la percentuale di utenti appartenenti alla fascia d'età più bassa risulti



molto più elevata, arrivando fino al 10%. Sono invece irrilevanti i numeri relativi a fasce superiori ai 34 anni; infatti, troviamo che soltanto il 2% dei domini totali presenta una maggioranza di utenti appartenenti a queste categorie.

Un ulteriore interessante confronto si effettua tra Europa e Stati Uniti, vediamo come i dati siano molto simili tra loro per quanto riguarda i siti a maggioranza utenti femminili, mentre si accentua ulteriormente la prevalenza di quelli a maggioranza maschili all'interno dei domini statunitensi. La spiegazione di questa differenza più elevata risiede nel fatto che negli Stati Uniti, la popolazione ha una prevalenza maschile al contrario dell'Europa.

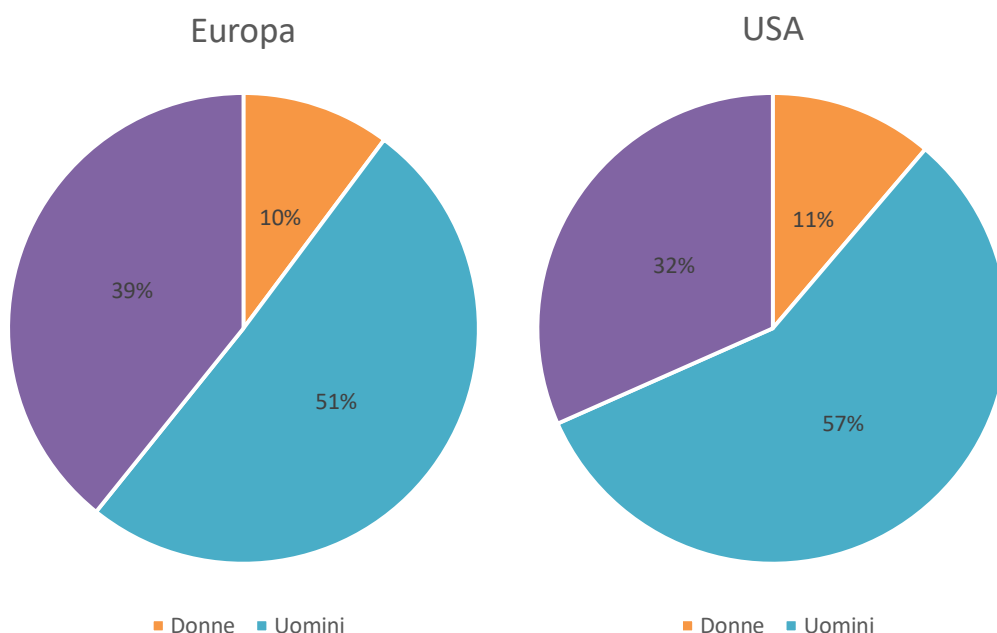


Tabella 15: suddivisioni per genere tra Europa e USA

Nel confronto a livello geografico delle fasce d'età, il dato più evidente è il fatto che a livello europeo troviamo soltanto tre fasce che vanno dai 18 ai 44 anni, mentre al di sopra i dati risultano irrilevanti. Negli Stati Uniti, benché, resti valida la prevalenza della fascia 25-34 anni, assumono importante rilevanza anche gli utenti, aventi più di 55 anni che corrispondono all'11% del totale. Questi dati sembrano evidenziare il fatto che la digitalizzazione ha raggiunto uno stato più avanzato negli USA, mentre nel vecchio continente risulta ancora molto limitata alla popolazione più giovane.

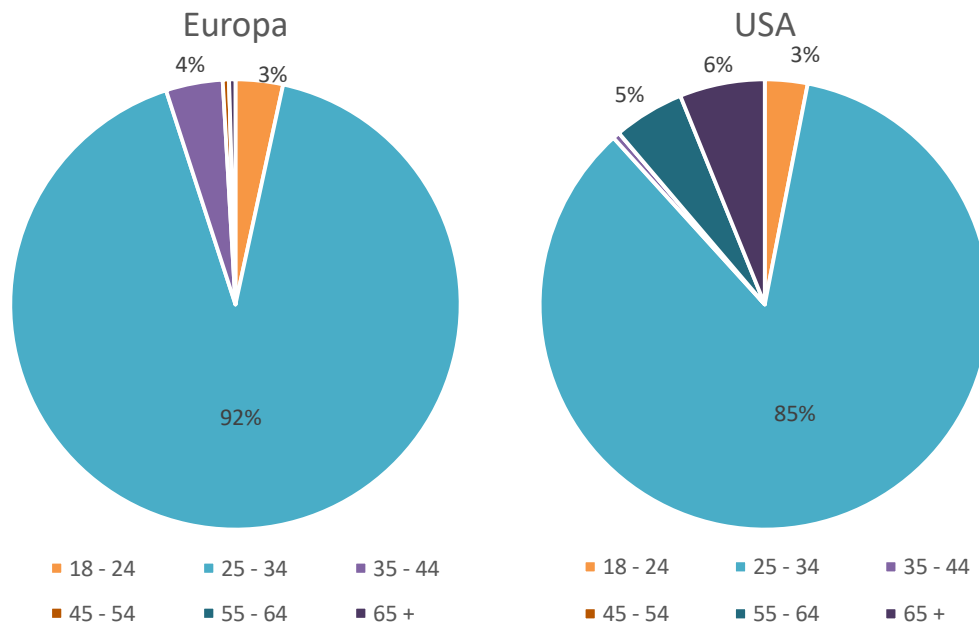


Tabella 16: suddivisioni anagrafiche tra Europa e USA

4.4 Statistiche temporali

Per introdurre l'analisi dei trend temporali, è necessario partire da una panoramica generale, per poi passare alle specifiche classi di studio.

Innanzitutto, i dati presentano quattro categorie principali:

- Il traffico, che permette di valutare il numero di accessi al dominio;
- La durata delle visite, che esprime quanto tempo l'utente resta all'interno del sito web;
- Il numero di pagine visitate all'interno del dominio, che rappresenta una misura del livello di iterazione tra l'utente e i contenuti;
- Il Bounce Rate, cioè la frequenza di rimbalzo, ovvero il numero di utenti in percentuale che chiudono il sito dopo essere entrati senza aver nessuna iterazione con esso.

I primi tre valori ovviamente vanno valutati in maniera crescente, cioè più è elevato il valore, maggiore è l'iterazione tra utente e dominio, mentre per l'ultimo indicatore bisogna valutare l'inverso; infatti, tutti i siti auspicano un valore più basso possibile.



Di seguito vengono riportati i confronti temporali tra Europa e Stati Uniti a livello generale, cioè comprendenti in aggregato tutte le classi in esame, che verranno poi dettagliate successivamente.

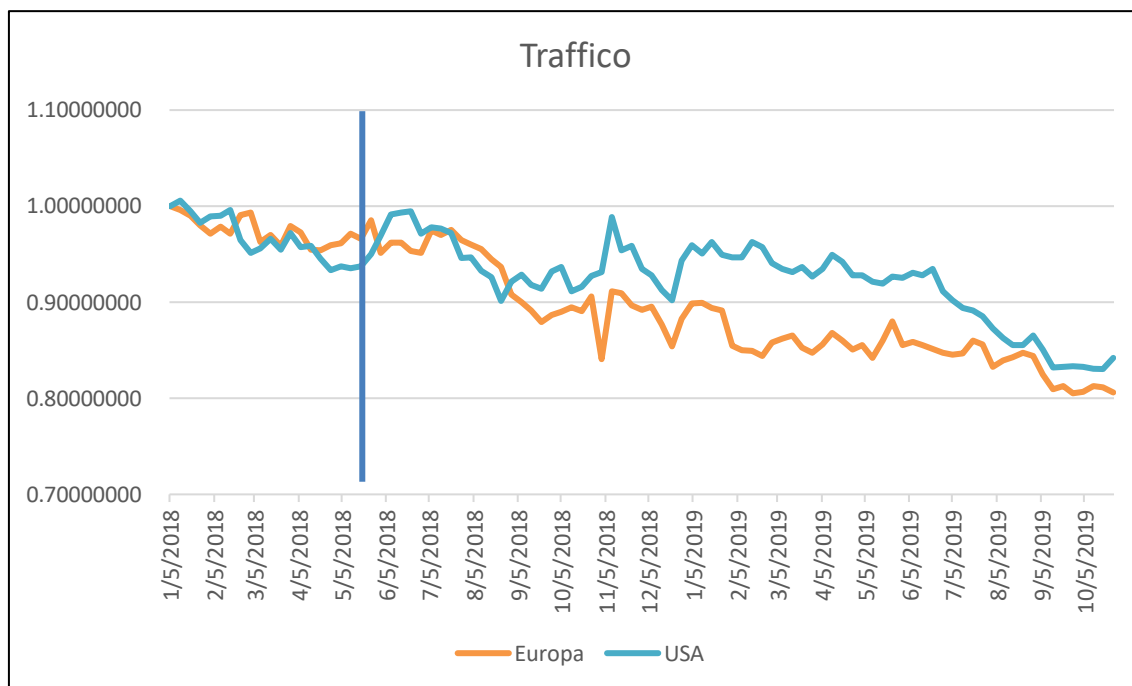


Tabella 17: confronto traffico medio tra Europa e USA

Da questo grafico emerge come successivamente all'emissione del GDPR, avvenuta nel maggio del 2018 i valori di traffico europei e statunitensi, che prima erano sovrapposti, subiscano entrambi una discesa, che però risulta più accentuata a livello europeo; infatti, dall'agosto 2018 in poi il traffico del vecchio continente risulta sempre al di sotto di quello degli USA di un valore pari a circa l'8%. Questo dimostra come il regolamento europeo abbia avuto un forte impatto sui cittadini dell'Unione, portandoli a ridurre di un valore medio del 10% circa il loro traffico totale, riduzione che non è avvenuta nel continente americano, anche se troviamo una lieve discesa anche qui, di circa il 4% in media.

Un dato interessante emerge in corrispondenza di luglio del 2019, dove anche per gli Stati Uniti si registra una forte discesa del traffico medio, che porta i valori poco sopra quelli europei.

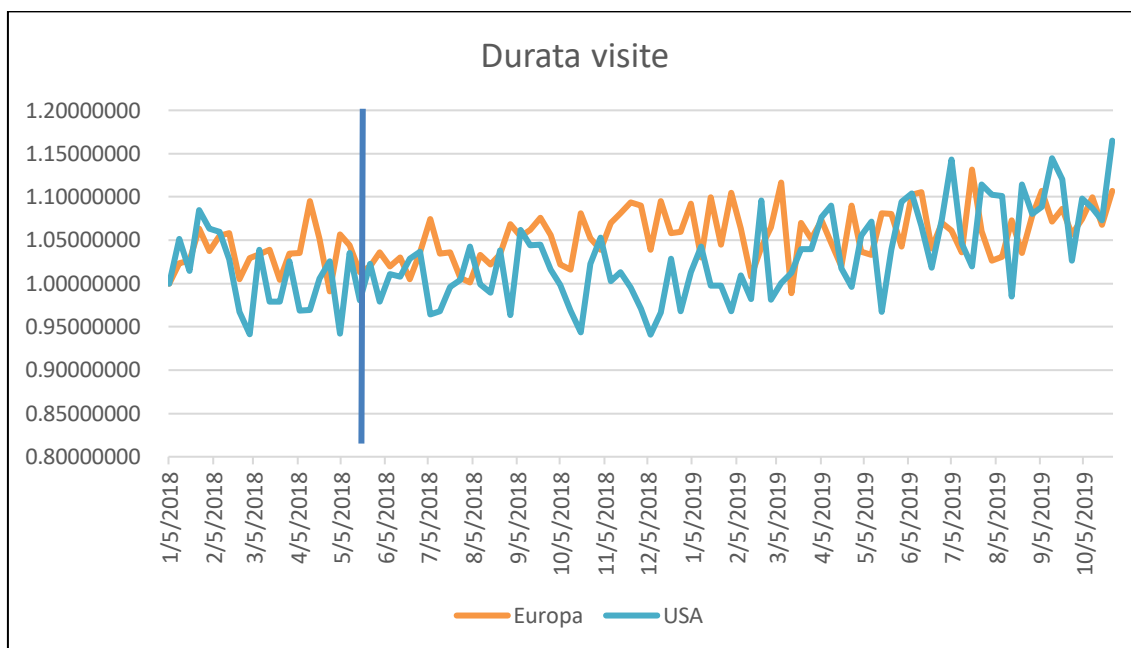


Tabella 18: confronto durata visite tra Europa e USA

Per quanto riguarda la durata delle visite non emergono variazioni significative in seguito all'entrata in vigore, anzi i valori europei si mantengono in media più elevati, anche se si registra una crescita media dei dati americani.

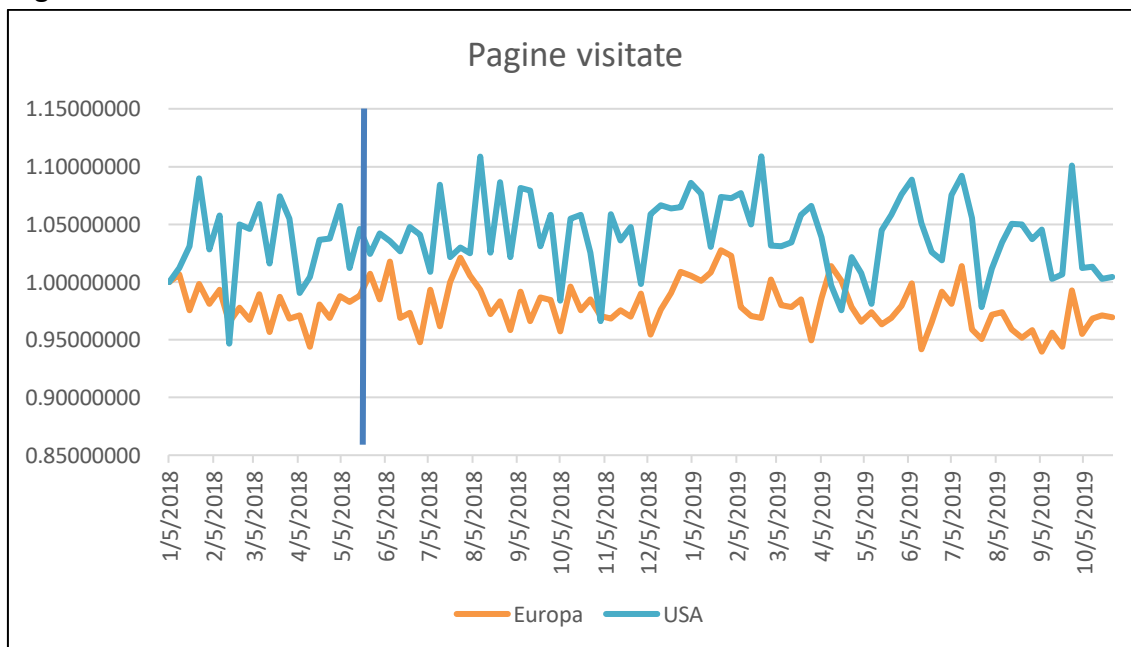


Tabella 19: confronto pagine visitate tra Europa e USA



Anche in questo caso non si vedono variazioni degne di nota in quanto i valori mantengono un trend costante, è interessante notare come i valori dei paesi UE siano in media il 10% inferiori ai valori degli USA, nonostante la durata visite fosse superiore, questo mostra come gli utenti europei abbiano una tendenza a analizzare per un tempo maggiore i contenuti, ma soffermandosi solo su alcuni di essi, mentre a livello statunitense si predilige una panoramica di più argomenti, meno dettagliati.

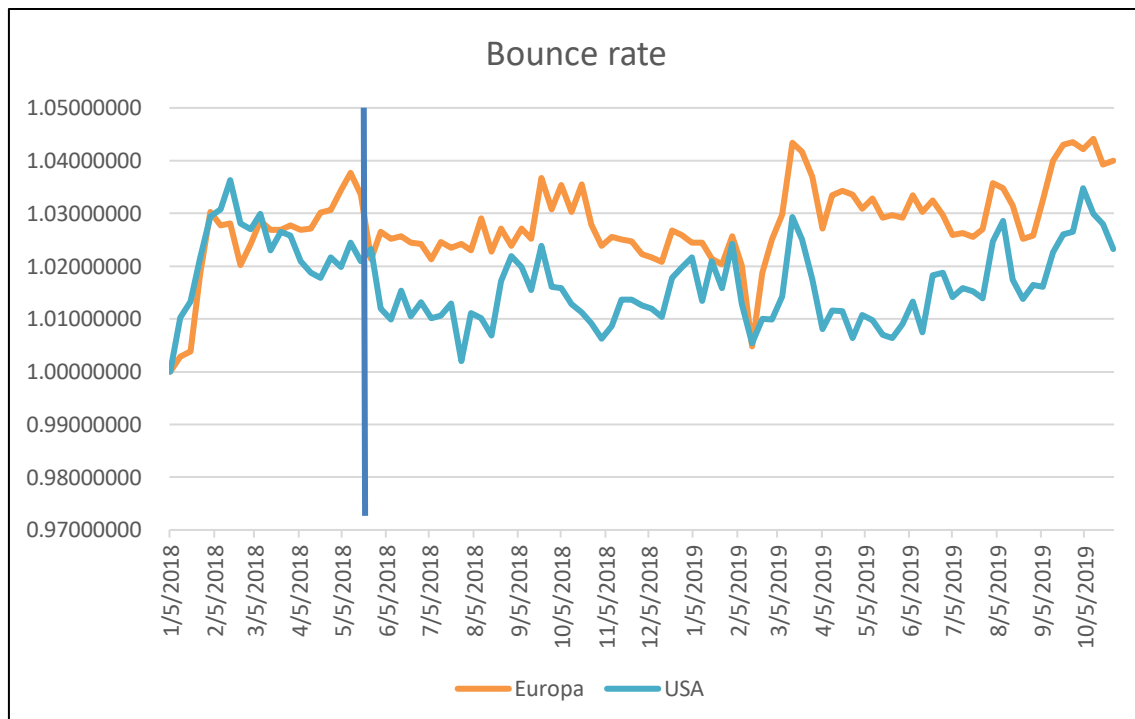


Tabella 20: confronto Bounce Rate tra Europa e USA

Quest'ultimo indicatore riporta un dato molto rilevante ai fini dello studio, infatti è visibile che a partire dal maggio 2018 (data in cui è entrato in vigore il GDPR) i valori del Bounce Rate, che precedentemente risultavano molto simili, abbiano intrapreso tendenze diverse, assestandosi su una differenza media di circa il 2%, il dato europeo risulta ovviamente più elevato e questo sembra dimostrare come il regolamento introdotto abbia modificato il comportamento degli utenti soggetti ad esso.

Infatti, se prima i dati erano simili e quasi sovrapposti, in seguito si vede come gli utenti europei siano maggiormente portati a uscire dal sito senza interagire con esso, questo potrebbe trovare la causa nell'informativa relativa alla privacy che tutti i domini devono presentare nella loro landing page, chiedendo a chi interagisce con essi di accettarla per



poter proseguire. Molti utenti leggendo l'informativa potrebbero a questo punto decidere di interrompere la loro iterazione.

Da questa analisi preliminare è emerso come gli unici valori influenzati dal nuovo regolamento siano il traffico e il bounce rate, mentre la durata delle visite e il numero di pagine visitate non subiscono variazioni significative. Tale tendenza risulta evidente anche all'interno delle singole classi di analisi citate, per questo nel proseguo del documento verranno prese in considerazione soltanto statistiche relative ai primi due indicatori citati sopra.

4.4.1 Statistiche di confronto: Media e Social

Per svolgere questo primo confronto puntuale è necessario suddividere i siti nelle due categorie presenti nel dataset, e per ognuna di esse effettuare un confronto tra i valori statunitensi e quelli europei.

Partendo dai valori relativi ai Media:

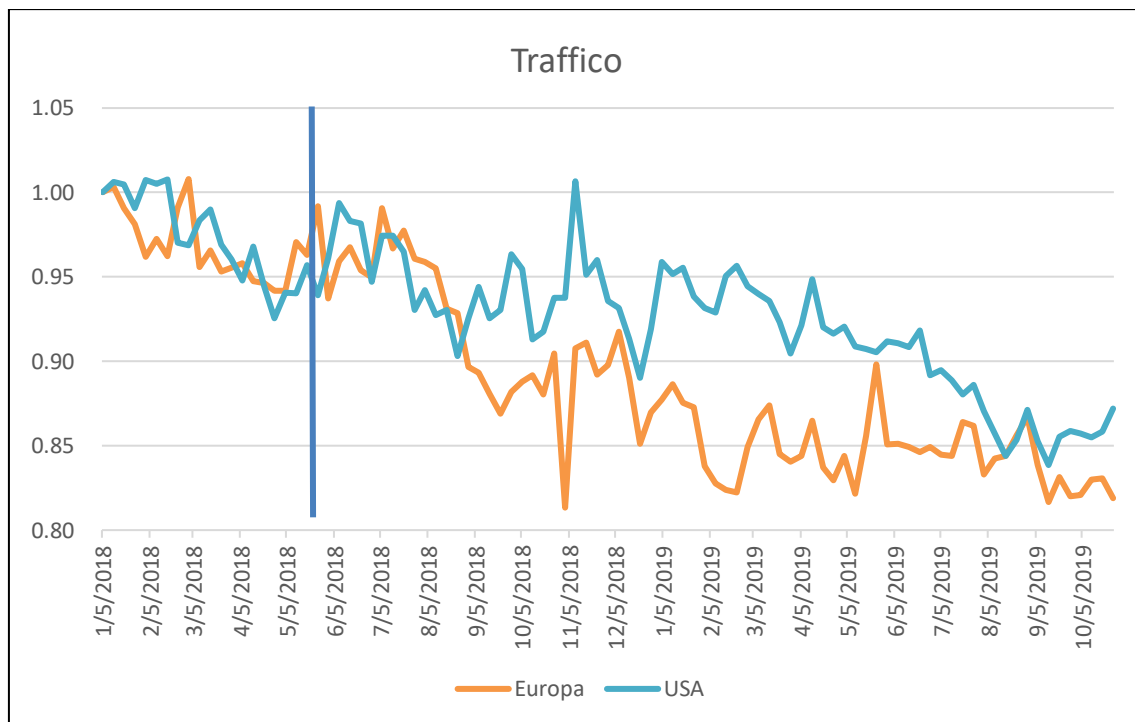


Tabella 21: Confronto traffico medio Media tra Europa e USA

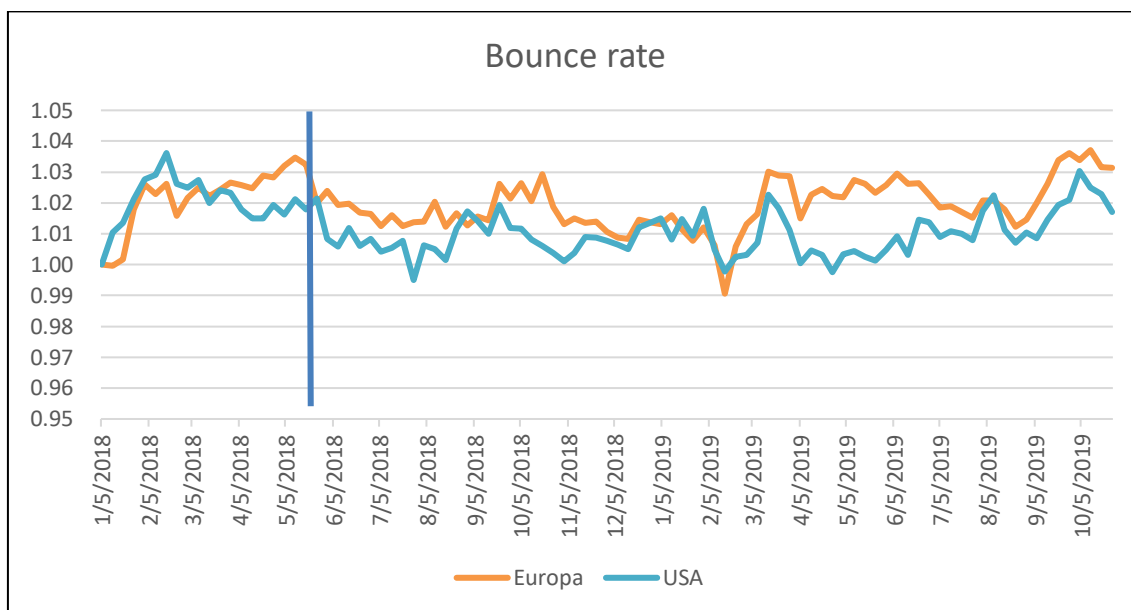


Tabella 22: Confronto bounce rate medio Media tra Europa e USA

Entrambi i valori sembrano dimostrare i risultati ottenuti nel confronto generale, con un traffico europeo che decresce rispetto a quello statunitense in seguito al GDPR e un bounce rate che aumenta.

Relativamente ai Social:

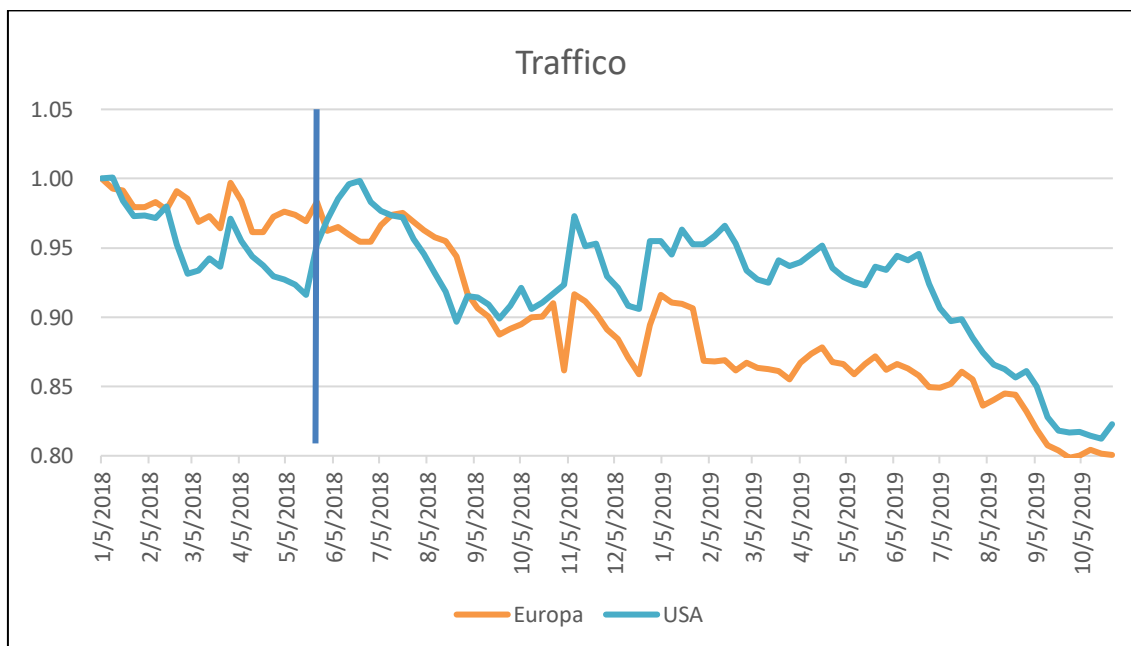


Tabella 23: confronto traffico medio Social tra Europa e USA

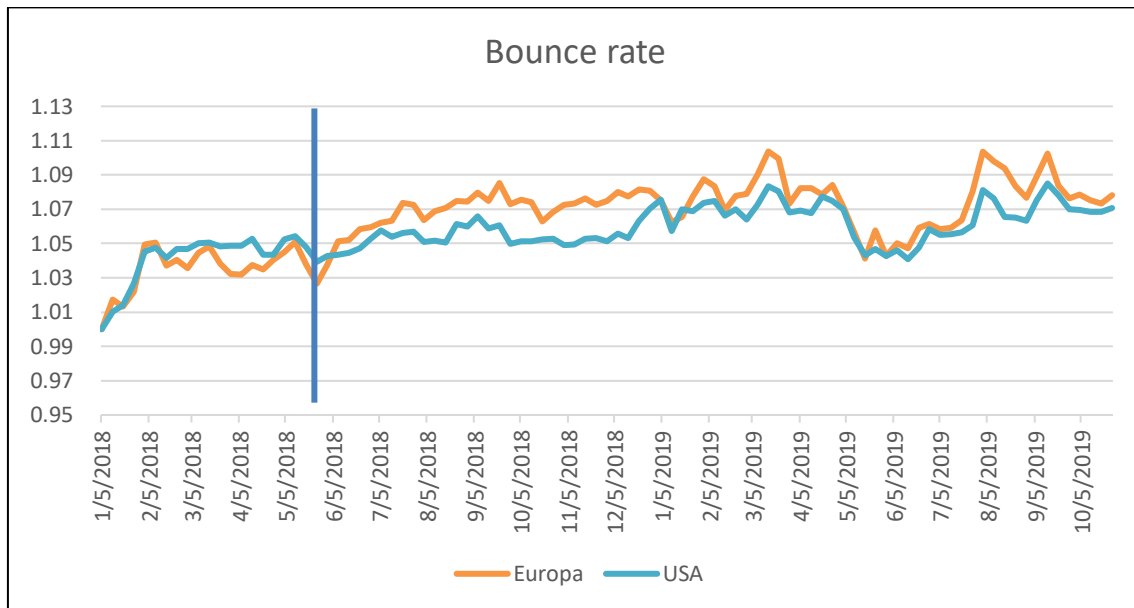


Tabella 24: confronto bounce rate medio Social tra Europa e USA

Come per i media il traffico medio europeo in seguito al GDPR si distanzia da quello degli Stati Uniti, negativamente, e al contempo il bounce rate aumenta in percentuale maggiore in Europa.

In conclusione, si può affermare che non si riscontrano differenze significative nel comportamento degli utenti nell'approcciarsi ai siti di News rispetto che ai Social, infatti per entrambe le categorie i trend risultano molto simili e confermano i risultati rilevati a livello generale.

4.4.2 Statistiche relative al confronto per genere

Il primo confronto di questa categoria di studi è effettuato comparando i valori per gli uomini e le donne europei, calcolati utilizzando il differenziale per ciascun genere tra il valore europeo e il corrispondente valore statunitense, in modo da ottenere risultati confrontabili e imputabili direttamente al GDPR, per valutare se ci siano differenze significative.

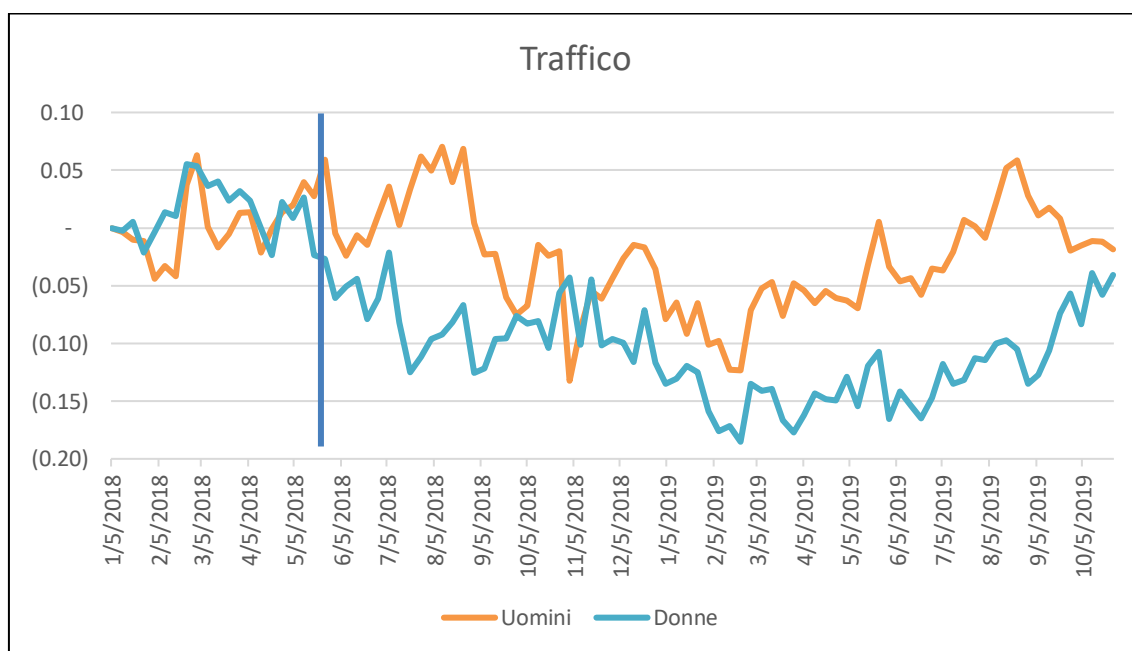


Tabella 25: confronto traffico medio tra uomini e donne europei

Da questa analisi si evidenzia una discesa in seguito all'emissione del regolamento, che risulta più marcata per il genere femminile; infatti, se prima del maggio 2018 la differenza era molto ridotta, successivamente, troviamo una differenza di quasi il 20% per alcuni mesi, che si assesta poi intorno al 15% nel dicembre dello stesso anno.

Questo dato sembra mostrare come le donne siano state maggiormente influenzate, soprattutto nel primo periodo, con una riduzione significativa del loro traffico medio, pari a circa il 25-30% rispetto all'inizio dell'anno. Per il genere maschile questa riduzione, pur essendo presente è inferiore e si assesta intorno al 12%.

Inoltre, l'effetto del GDPR si mostra subito in ambito femminile; infatti, la discesa più rilevante avviene a partire dal luglio del 2018, quella maschile invece un paio di mesi dopo.

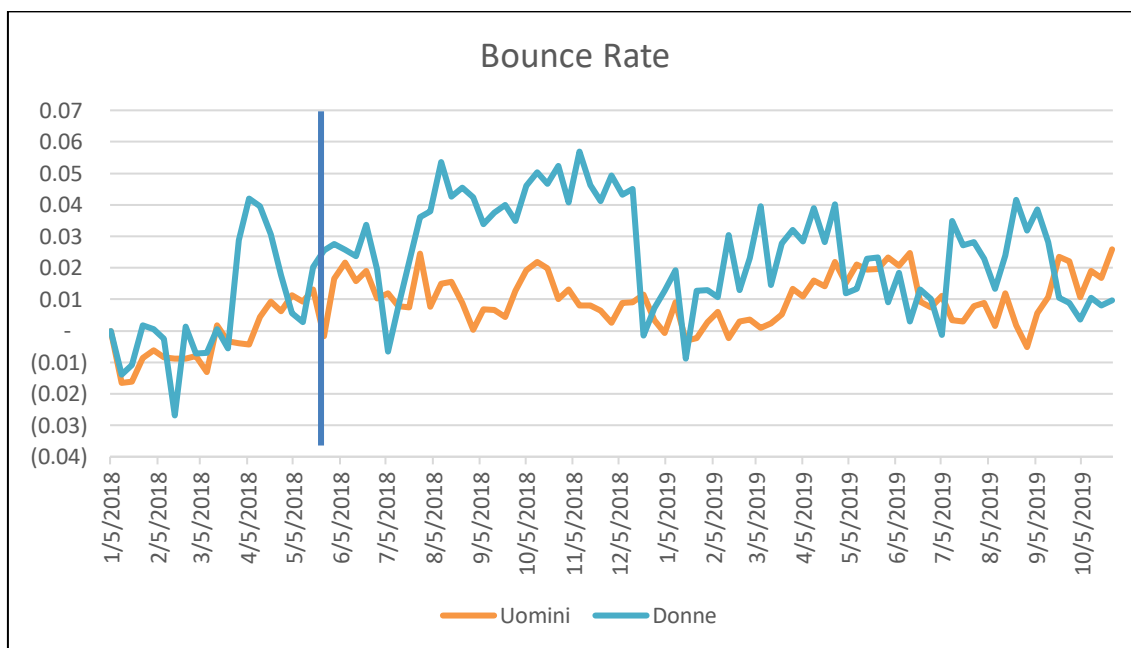


Tabella 26: confronto bounce rate tra uomini e donne europei

Anche per questo secondo indicatore si manifesta la tendenza descritta precedentemente, con un impatto molto più elevato sulle donne, si vede come prima dell'emissione del regolamento i valori tendevano a presentare un bounce rate molto più elevato per gli uomini, mentre successivamente i valori tendono a equipararsi, arrivando nel lungo periodo, in seguito all'emissione della prima grande sanzione citata precedentemente, a presentare numeri addirittura maggiori per il genere femminile. La crescita dei valori si manifesta per entrambi i generi: nelle donne arriva a toccare il 6%, negli uomini il 5%.



Procedendo nell'analisi è utile confrontare gli indicatori per entrambi i generi separatamente, valutando le differenze con il continente americano.

Partendo dagli uomini troviamo i due seguenti grafici:

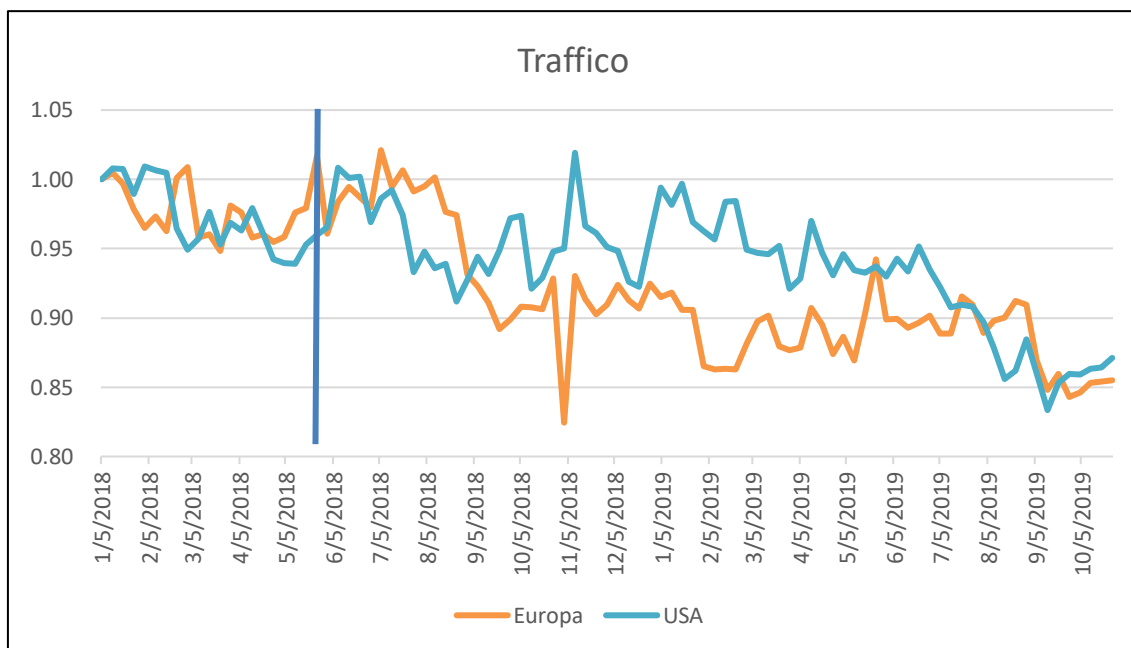


Tabella 27: confronto traffico medio degli uomini tra Europa e USA

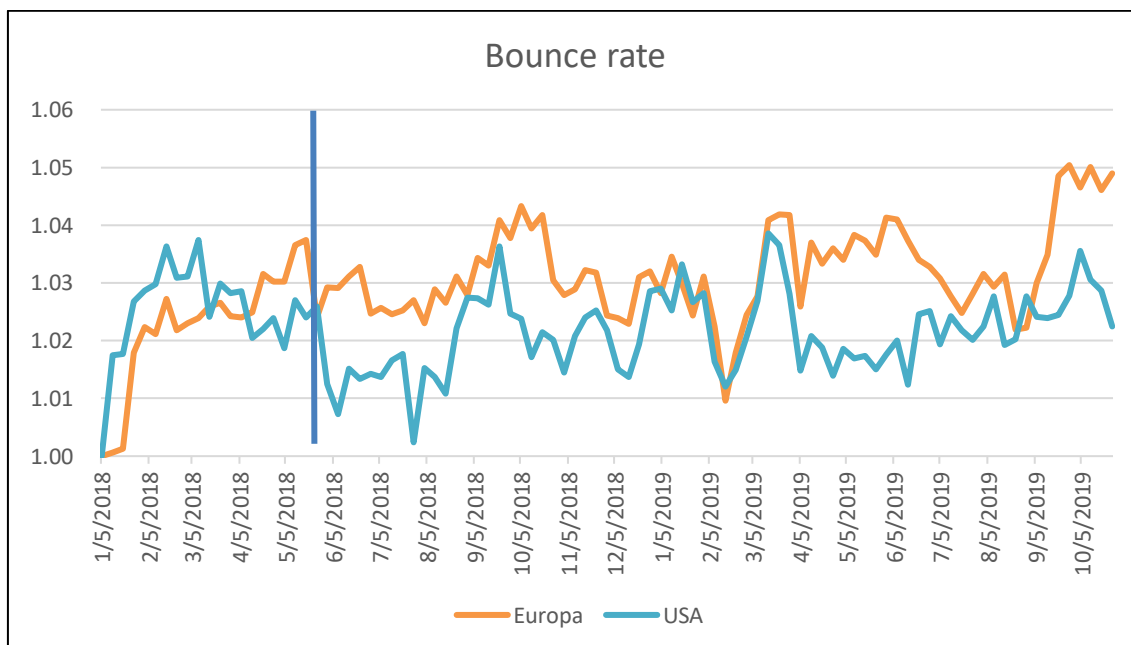


Tabella 28: confronto Bounce Rate degli uomini tra Europa e USA



Entrambi confermano le tendenze descritte in precedenza, è visibile che per quanto riguarda il traffico i valori europei in seguito all'entrata in vigore subiscono una forte riduzione, mentre quelli americani restano costanti, questo porta un passaggio da dati simili, sovrapponibili, a dati che differiscono di circa il 5%.

Anche il Bounce Rate europeo si alza rispetto a quello americano, ma in questo caso la differenza si attesta su valori minori, circa il 2%.

Il confronto tra i valori di Europa e Stati Uniti per quanto riguarda il genere femminile:

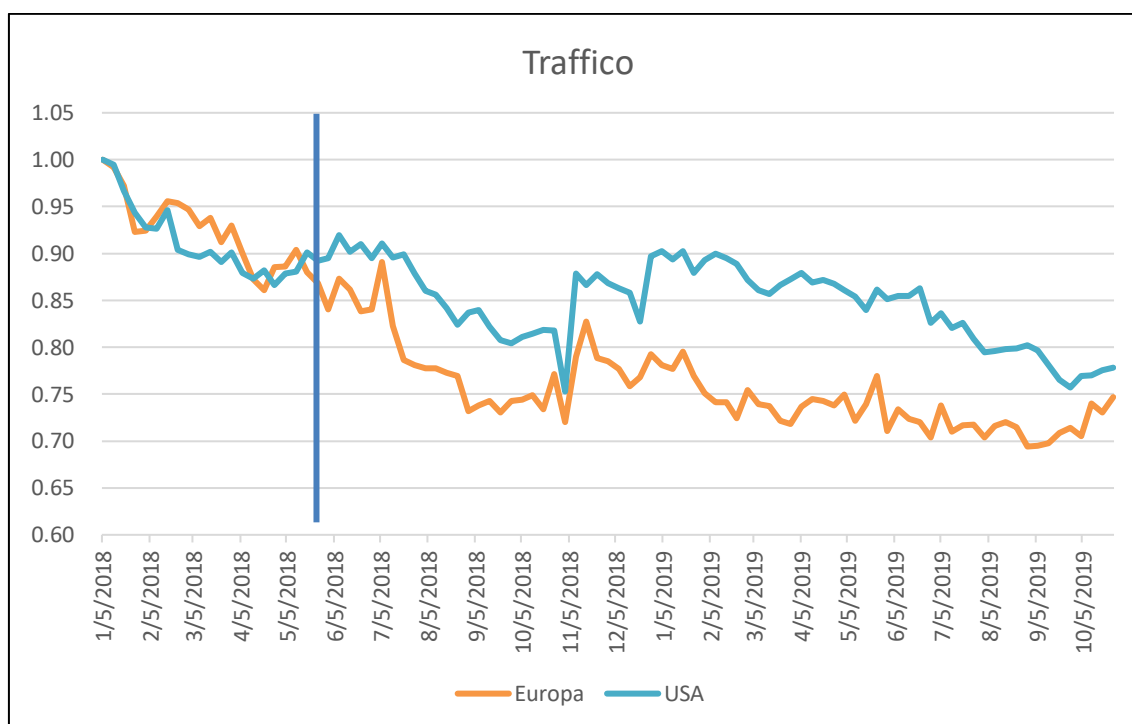


Tabella 29: confronto traffico medio delle donne tra Europa e USA

Trend molto simile per quanto riguarda il traffico medio a quello riscontrato per gli uomini, ma con valori molto più elevati: una differenza tra UE e USA successiva al GDPR che risulta in media pari al 10%.

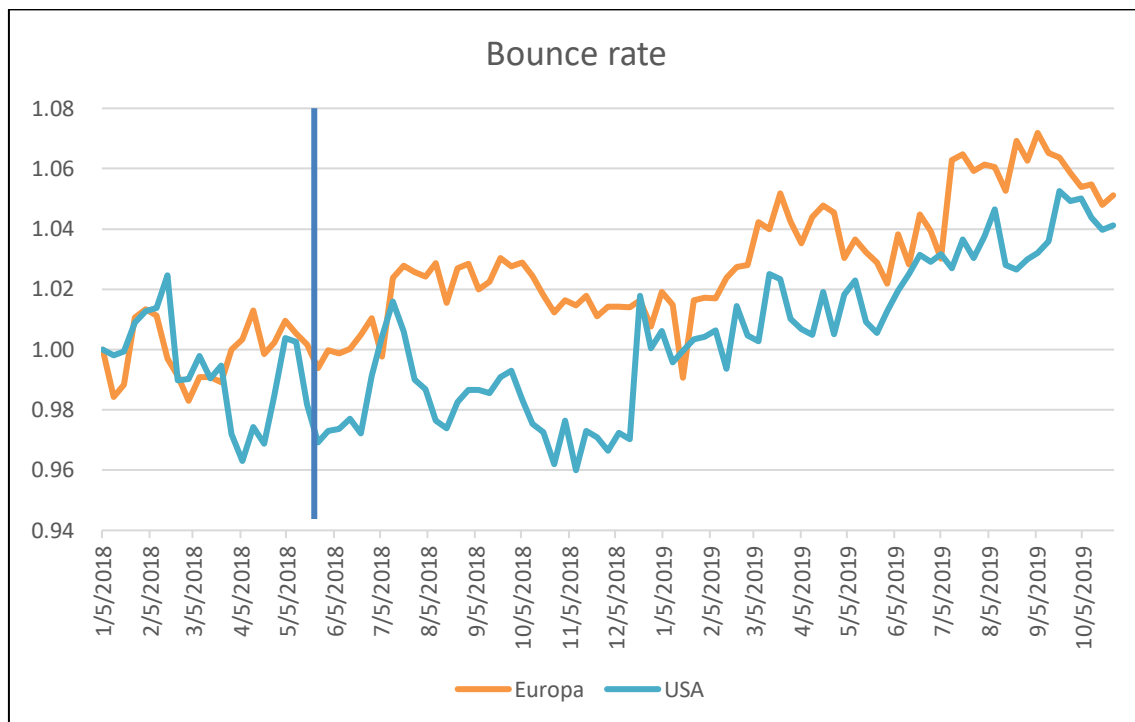


Tabella 30: confronto Bounce Rate delle donne tra Europa e USA

In questo caso la differenziazione risulta nettamente inferiore, se non in un periodo di circa 5 mesi successivi all'emissione, ma nel lungo periodo i valori tornano a essere molto simili, con una distanza comparabile a quella evidenziata nel genere maschile che riporta un dato pari al 2% circa.

In conclusione, è possibile affermare come l'effetto del GDPR sia riscontrabile in entrambi i generi analizzati, con valori molto simili per quanto riguarda la frequenza di rimbalzo. Nel valutare il traffico medio al contrario si manifesta una certa eterogeneità, dato che in entrambi vediamo una forte riduzione nella comparazione con i dati americani, ma questa diminuzione si manifesta con un valore doppio per le donne rispetto agli uomini.

Da questo si evince come l'effetto del banner informativo relativo alla privacy imposto dal regolamento europeo abbia avuto lo stesso effetto sia per i maschi sia per le femmine andando cioè ad aumentare il numero di utenti che escono dal dominio una volta visualizzata questa informativa. Al contrario l'effetto generale della nuova regolamentazione ha avuto un impatto più rilevante sul genere femminile dimostrando come questo sia più sensibile riguardo a temi di privacy e protezione dei dati personali.

4.4.3 Statistiche relative al confronto anagrafico

Anche per questa seconda analisi si parte da un confronto generico a livello europeo tra le fasce d'età citate sopra.

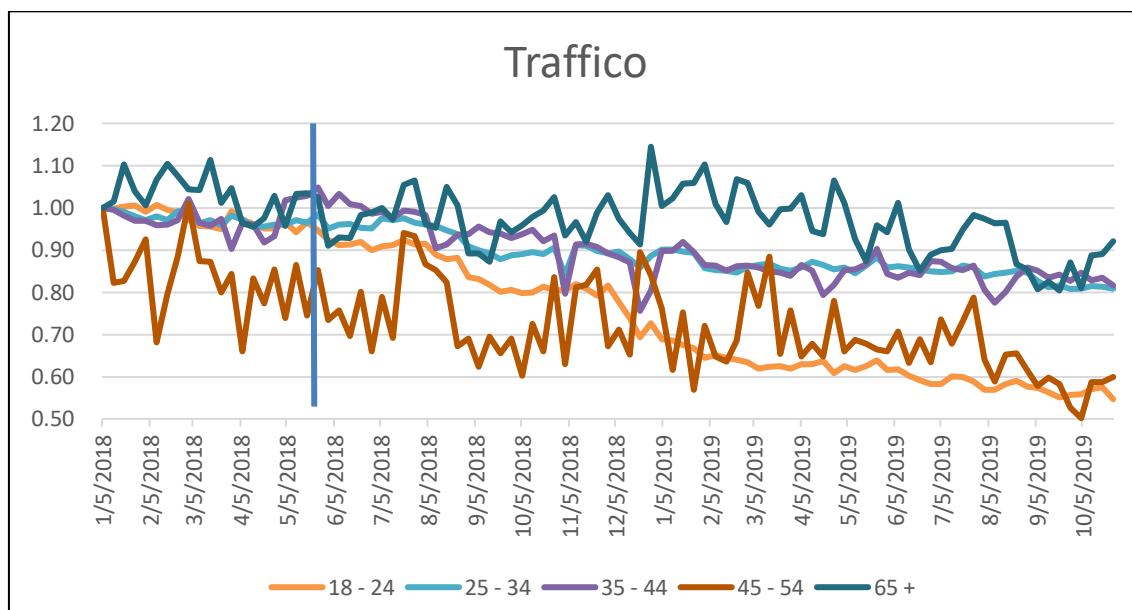


Tabella 31: confronto traffico tra le fasce d'età a livello europeo

Dal grafico emerge come nel lungo periodo tutte le fasce abbiano subito una riduzione del traffico medio, ma con valori nettamente differenti:

- Nella fascia tra i 18 e i 24 anni e tra 45 e 54 la riduzione è stata pari al 40% circa
- Tra i 25 e i 44 la riduzione ha un valore del 18% circa
- Sopra i 65 anni in media dell'8%, ma con addirittura periodi caratterizzati da un aumento del traffico

Vediamo quindi una grande eterogeneità tra i risultati emersi, soprattutto per la fascia più anziana di popolazione, quest'ultimo dato può trovare spiegazione nel fatto che spesso le persone in questa categoria hanno meno familiarità con i temi della privacy e i pericoli legati ad essa che il web può nascondere.

Nell'intervallo tra i 45 e i 54 anni si notano oscillazioni di grande portata, con periodo in cui il traffico sembra risalire a valori antecedenti all'entrata in vigore del regolamento europeo.

Per tutte le classi si evidenzia come la riduzione maggiore di traffico avvenga a partire dal luglio 2018, un paio di mesi dopo l'entrata in vigore del GDPR

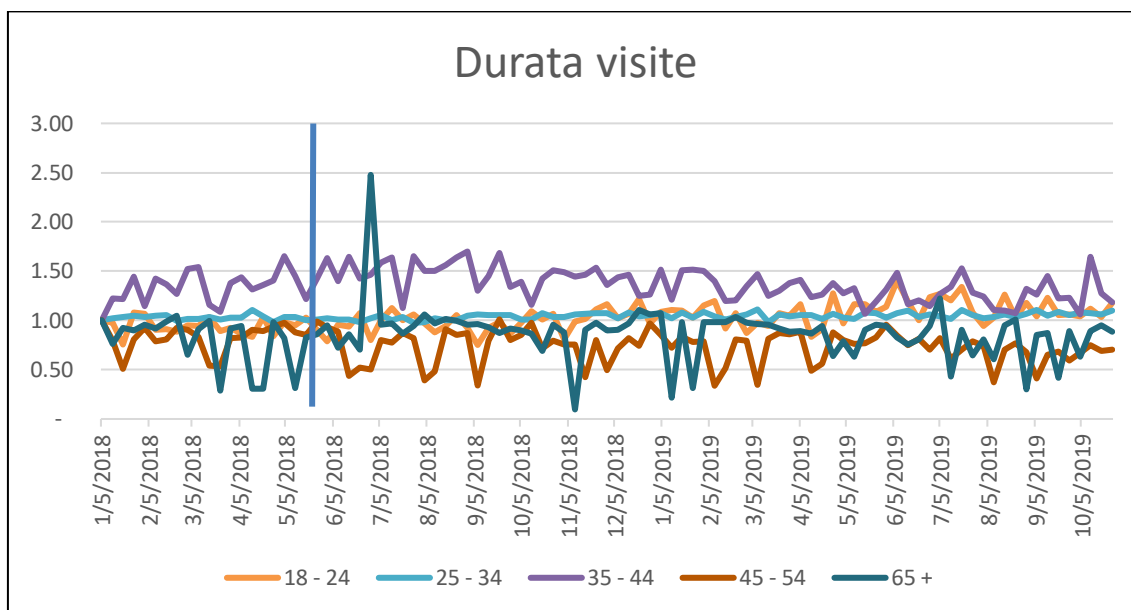


Tabella 32: confronto durata visite tra fasce d'età a livello europeo

Riguardo alla durata media delle visite si nota come per la categoria tra i 35 e i 44 anni presenti valori mediamente più alti della media del 50% circa; i più giovani, invece presentano valori medi e soggetti a basse oscillazioni; ad ultimo al di sopra dei 45 anni i valori tendono a presentare ampie oscillazioni e un valor medio inferiore rispetto alle altre fasce di circa il 25%.

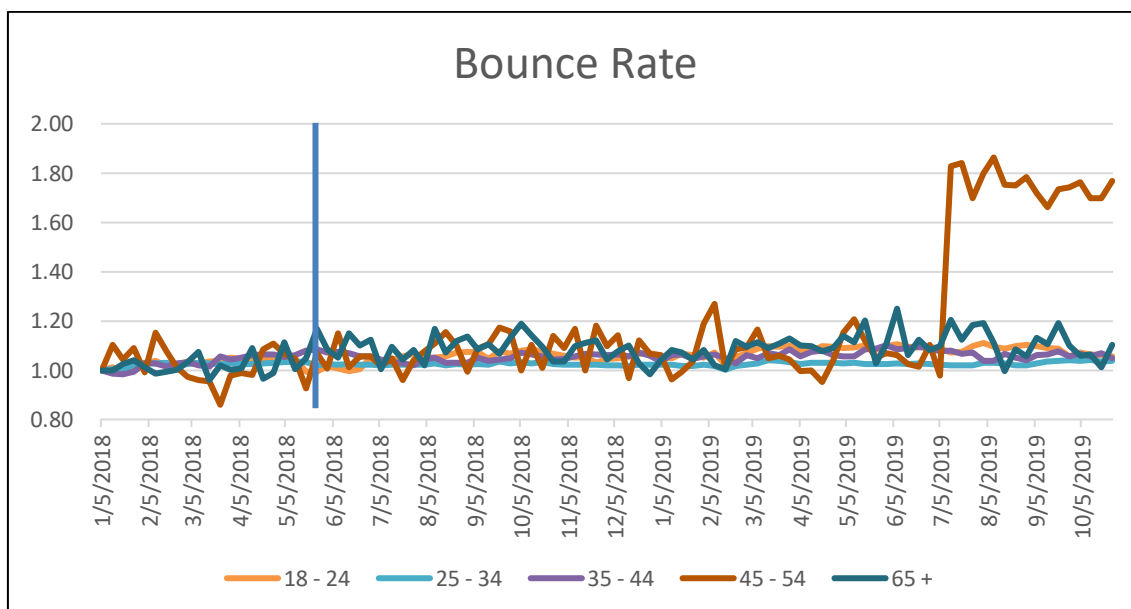


Tabella 33: confronto Bounce Rate tra le fasce d'età a livello europeo

Contrariamente all'indicatore precedente per quanto riguarda il Bounce Rate i valori sono simili ed equiparabili per tutte le classi, fatta eccezione per l'intervallo 45,54 anni dove a partire dal luglio 2019 si evidenzia un incremento molto elevato, pari a più dell'80%.

Tutte le altre classi sono caratterizzate da un aumento medio del bounce rate su valori simili compresi tra il 3 e il 10%. Con il valore maggiore raggiunto dagli over 65, una tendenza opposta a quella riscontrata nel traffico.

Tre classi, cioè quelle al di sotto dei 45 anni sono caratterizzate da una crescita costante, senza oscillazioni rilevanti, mentre al di sopra di quest'età la crescita risulta maggiore, ma caratterizzata da grandi oscillazioni, queste considerazioni dimostrano come l'uso di Internet sia molto più costante tra i più giovani, che fanno di esso un'abitudine e quindi mantengono valori simili, al netto della crescita portata dal GDPR.

Nelle seguenti pagine si procede a un confronto puntuale tra Europa e Stati Uniti per tutte le categorie per le quali sono presenti valori.

Partendo dalla fascia tra i 18 e i 24 anni:

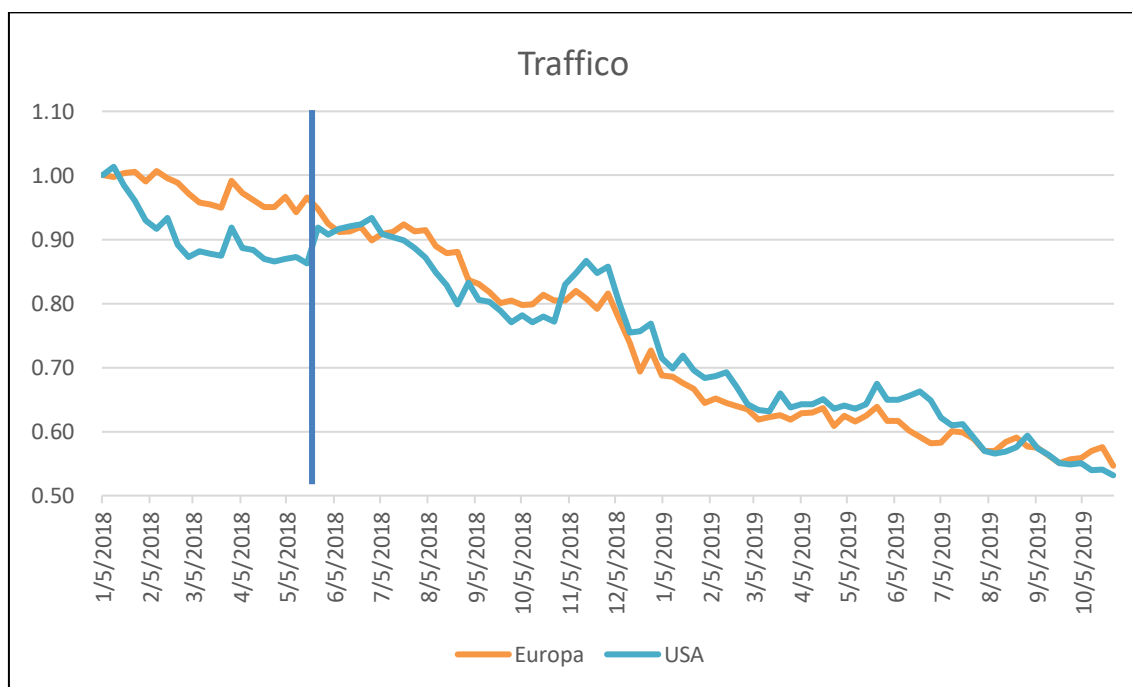


Tabella 34: confronto traffico medio 18-24 anni tra Europa e USA

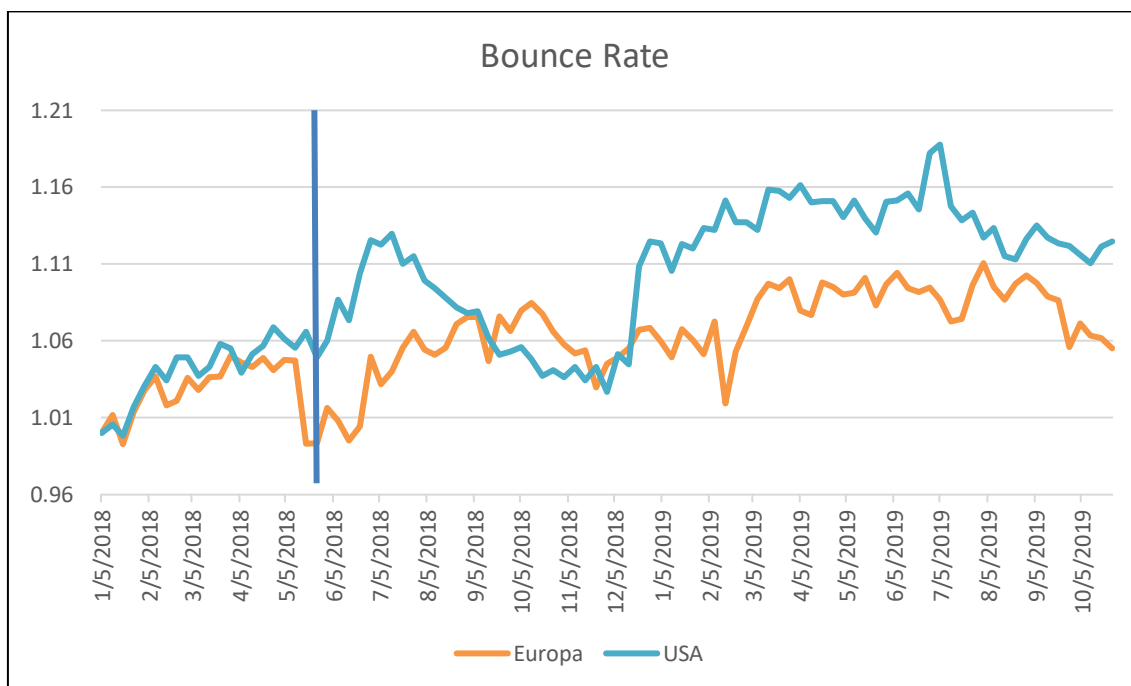


Tabella 35: confronto Bounce Rate 18-24 anni tra Europa e USA

Per quanto riguarda il traffico si vede una discesa costante sia a livello europeo sia a livello statunitense e quindi non imputabile al GDPR, nel Bounce Rate vediamo come risulta addirittura più alto quello americano e quindi anche in questo caso non si evidenziano impatti del nuovo regolamento europeo.



Politecnico
di Torino

Nella fascia tra 25 e 34:

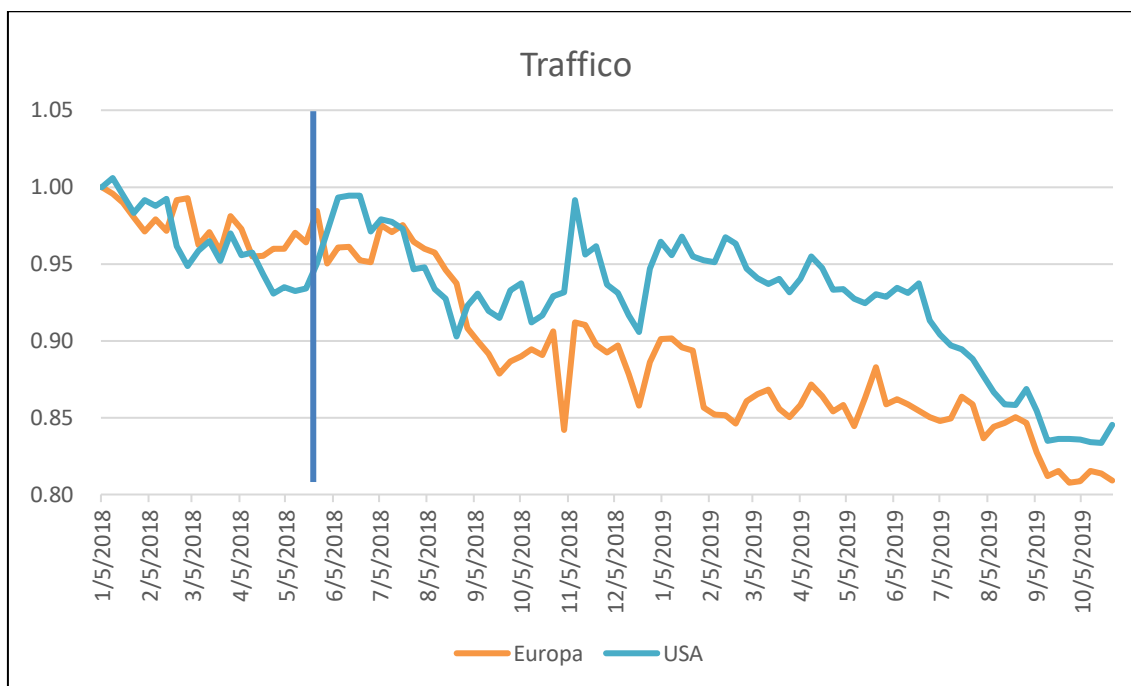


Tabella 36: confronto traffico medio 25-34 anni tra Europa e USA

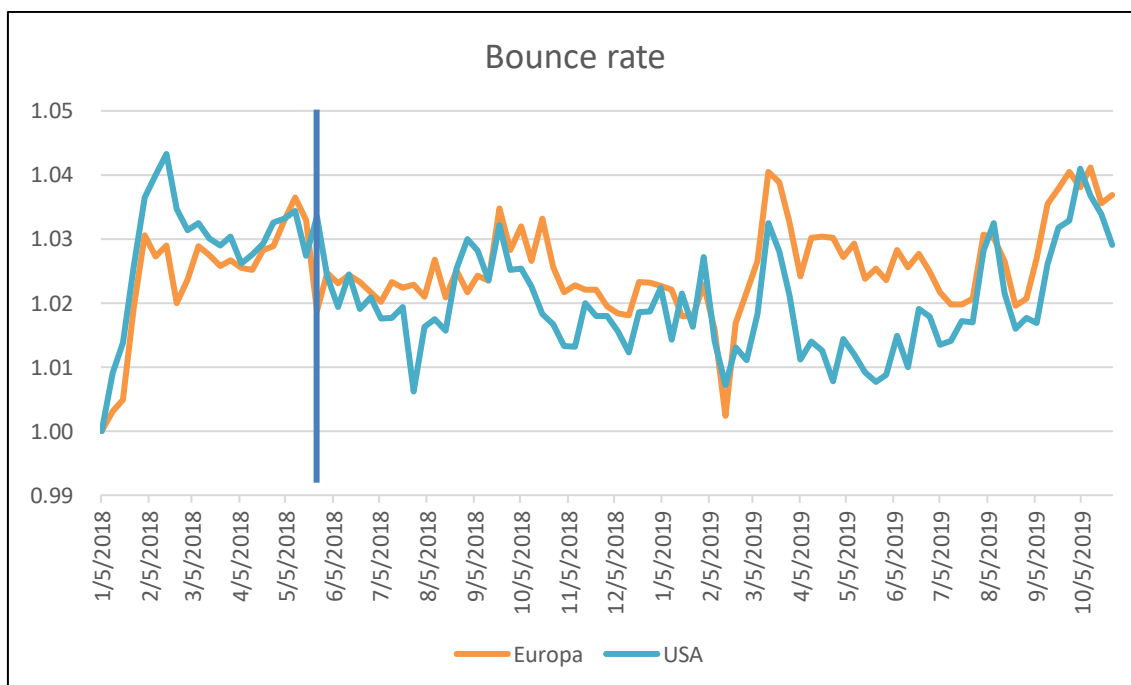


Tabella 37: confronto Bounce Rate 25-34 anni tra Europa e USA



In questo caso gli effetti del regolamento europeo risultano evidenti in quanto, per il traffico vediamo una riduzione netta di quello UE che partendo da valori sovrapponibili arriva a raggiungere un 5% in meno rispetto agli Stati Uniti, il bounce rate segue un andamento simile, cioè si attesta su valori più alti seppur con percentuali inferiori. Interessante notare come il secondo indicatore abbiamo una volatilità maggiore e in alcune rilevazioni torni ad assumere numeri simili tra le due aree geografiche.

Tra i 35 e i 44:

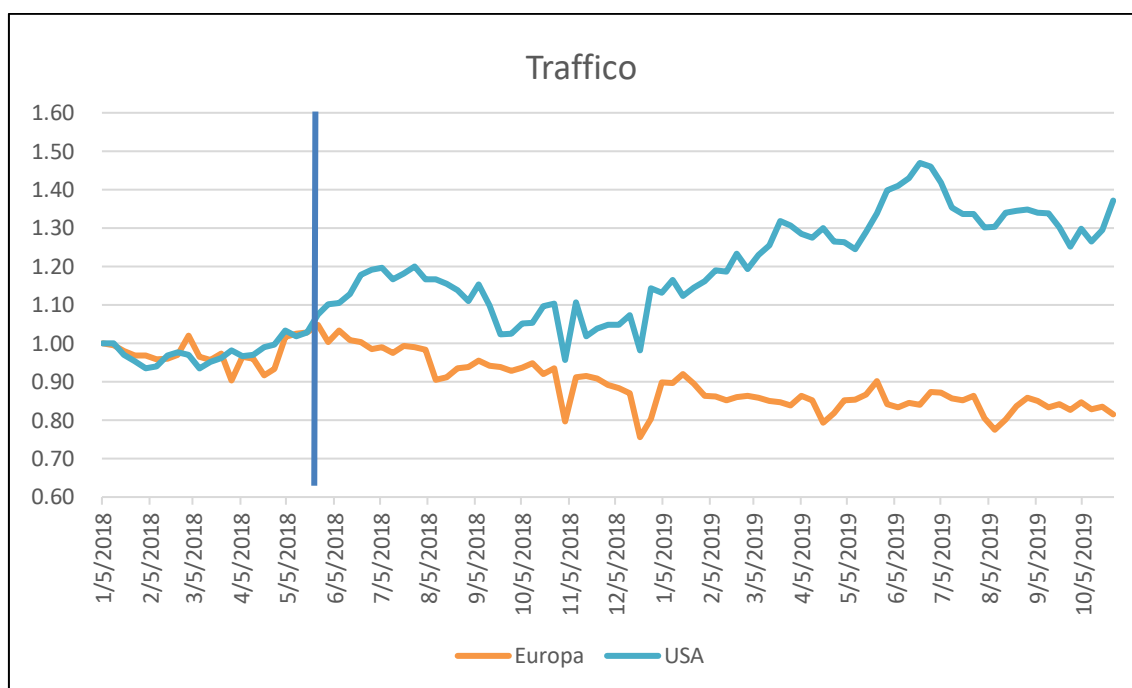


Tabella 38: confronto traffico medio 35-44 anni tra Europa e USA

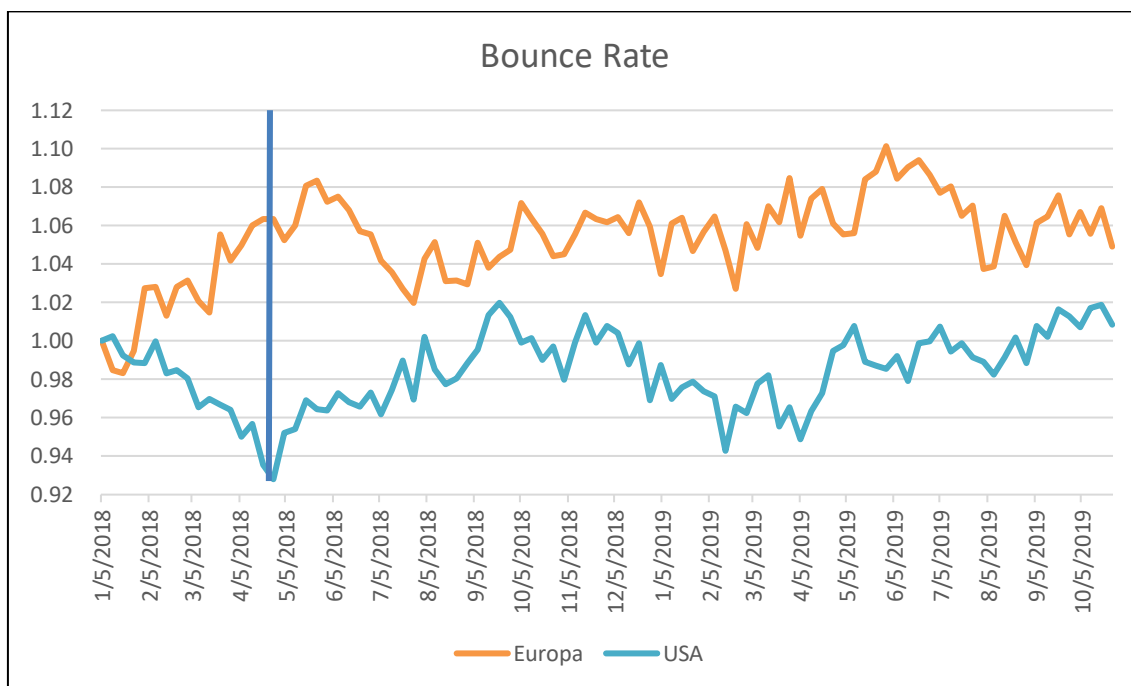


Tabella 39: confronto Bounce Rate 35-44 anni tra Europa e USA

Anche in questo caso i risultati confermano le teorie espone in precedenza con un traffico medio europeo che si attesta un 30% in meno in media rispetto a quello statunitense in seguito all'emissione del GDPR.

Nel secondo indicatore invece non si manifestano tendenze rilevanti in seguito all'entrata in vigore, infatti, risulta più elevato in Europa già a partire dal marzo 2018, questo dato potrebbe dipendere dall'annuncio della futura entrata in vigore, ma anche da fattori esterni sconosciuti e quindi non permette di trarre conclusioni a riguardo.



Ad ultima la fascia d'età over 65:

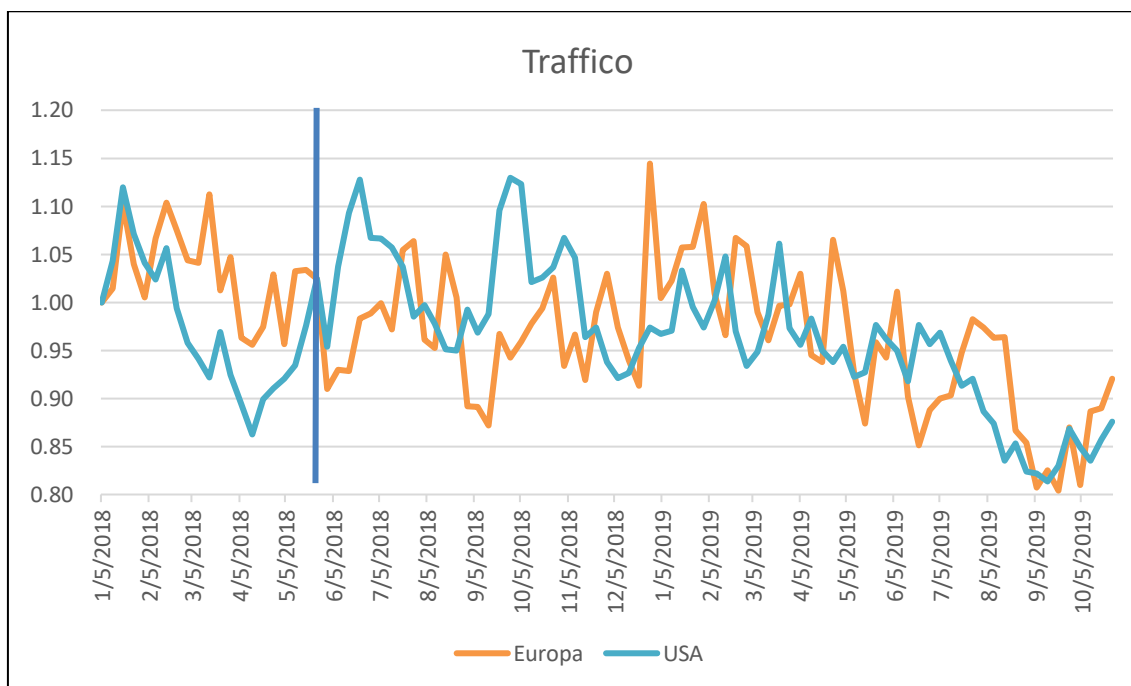


Tabella 40: confronto traffico medio 65+ anni tra Europa e USA

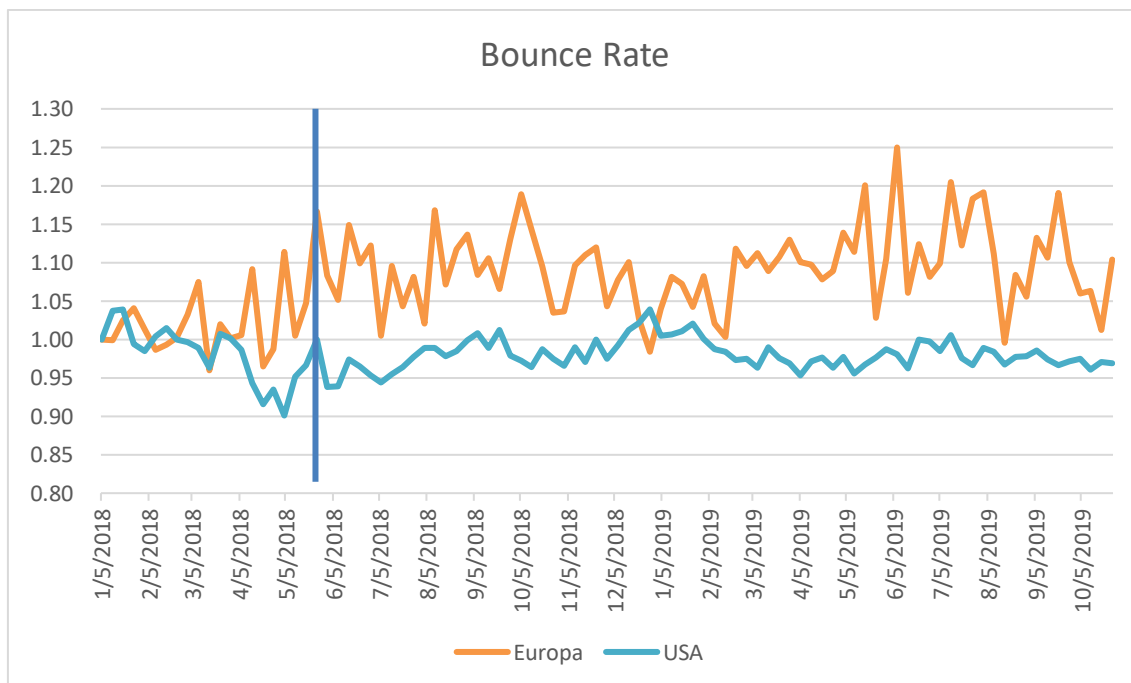


Tabella 41: confronto Bounce Rate 65+ anni tra Europa e USA



Dai dati di traffico non si possono estrapolare differenze significative a conferma della tesi descritta precedentemente, in quanto il traffico europeo subisce un decremento in corrispondenza del maggio 2018, ma da novembre dello stesso anno la tendenza sembra invertirsi, portando questo a superare i valori degli Stati Uniti.

Al contrario per l'indicatore relativo alla frequenza di rimbalzo troviamo conferma della tesi esposta in quanto è visibile come a partire dalla data di entrata in vigore del GDPR i valori europei si alzano in media del 15% mentre quelli statunitensi rimangono pressappoco costanti.

Dall'analisi effettuata possiamo concludere che ci troviamo di fronte a una certa eterogeneità per quanto riguarda gli effetti del regolamento europeo sulle diverse fasce d'età in esame. Coloro che sono stati influenzati in misura minore sono sicuramente i giovani al di sotto dei 24 anni, il cui comportamento non risulta cambiato per effetto del GDPR, la spiegazione a questo può essere trovata nella scarsa informazione e conoscenza delle tematiche relative alla privacy all'interno di questa fascia d'età, inoltre i più giovani tendono per la maggior parte a considerare i propri dati personali meno importanti, e quindi a tutelarsi meno. Un'ultima motivazione può essere il fatto che questo tipo di utenti è poco propenso a legger l'informativa sulla privacy che ogni dominio deve rendere disponibile e tende quindi ad accettarle senza soffermarsi. Al contrario, le altre fasce prestano attenzione a quest'informativa come dimostra l'aumento netto della frequenza di rimbalzo, valido per tutte le categorie rimanenti. I più anziani però come i più giovani non hanno ridotto il loro traffico in virtù del GDPR.



5. Modelli econometrici

È stata valutata un'equazione del tipo:

$$y_{D,t,c} = \beta_0 + \beta_1 post_t + \beta_2 treat_c + \beta_3(post \times treat)_{t,c} \\ + \beta_4(gender \times post \times treat)_{t,c} + \beta_5(age \times post \times treat)_{t,c} + \epsilon$$

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

- Traffico (tramite trasformazione logaritmica);
- Durata visite (tramite trasformazione logaritmica);
- Numero 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 (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;
- **gender x post x treat** rappresenta l'iterazione tra la precedente e la variabile di genere, assume valore pari a 1 qualora la visita sia europea, successiva al GDPR e relativa a un dominio a prevalenza utenti uomini;
- **age x post x treat** rappresenta l'iterazione tra la $post \times treat$ e la variabile di età, assume valore pari a 1 qualora la visita sia europea, successiva al GDPR e relativa a un dominio a prevalenza utenti under 35;
- 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 sei 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 tempo;



- **Specificazione 3:** valuta gli effetti fissi su dominio, tempo e provenienza del sito di ricerca (country);
- **Specificazione 4:** per valutare se ci siano stati degli effetti eterogenei tra i generi degli utenti;
- **Specificazione 5:** per valutare se ci siano stati degli effetti eterogenei tra utenti over e under 35.

Di seguito i risultati ottenuti:

VARIABLES	(1) log_traffic	(2) log_traffic	(3) log_traffic	(4) log_traffic	(5) log_traffic	(6) log_traffic
postxtreat	-0.0506 (0.0830)	-0.0613 (0.0495)	-0.0240 (0.0362)	-0.00903 (0.0368)	0.00618 (0.529)	0.171 (0.124)
treat	-1.241*** (0.135)	-2.270*** (0.202)	-2.294*** (0.203)			
post	-0.0937 (0.0733)	-0.0230 (0.0375)				
genderxpostxtreat					-0.0161 (0.533)	
agexpostxtreat						-0.190 (0.123)
Constant	14.31*** (0.117)	14.96*** (0.134)	15.00*** (0.136)	12.50*** (0.175)	12.50*** (0.197)	12.50*** (0.175)
Observations	38,096	38,096	38,096	38,096	38,096	38,096
R-squared	0.187	0.764	0.765	0.835	0.835	0.835
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1						

Tabella 42: regressione lineare relativa al traffico

Relativamente al traffico si può notare come in nessuna delle variabili di studio emergano coefficienti statisticamente significativi, contrariamente ai risultati ottenuti nei confronti temporali svolti nel capitolo precedente tramite la piattaforma Excel. In particolare:

- Dalla prima riga non emergono differenze significative a livello di categorie aggregate di siti Media e Social;
- La quarta riga rappresenta che non vi sono differenze di comportamenti tra utenti uomini e donne;
- La quinta riga indica omogeneità di comportamento tra le fasce d'età.



VARIABLES	(1) log_visit_ duration	(2) log_visit_ duration	(3) log_visit_ duration	(4) log_visit_ duration	(5) log_visit_ duration	(6) log_visit_ duration
postxtreat	-0.00140 (0.0656)	-0.0549 (0.0455)	-0.0576 (0.0392)	-0.0579 (0.0395)	-0.554*** (0.199)	-0.151** (0.0610)
treat	0.000132 (0.0830)	-0.0881 (0.116)	-0.0864 (0.115)			
post	-0.0487 (0.0575)	0.0447 (0.0383)				
genderxpostxtreat					0.525*** (0.193)	
agexpostxtreat						0.0978 (0.0596)
Constant	9.579*** (0.0744)	9.601*** (0.0784)	9.564*** (0.0818)	9.506*** (0.108)	9.587*** (0.119)	9.505*** (0.108)
Observations	38,096	38,096	38,096	38,096	38,096	38,096
R-squared	0.001	0.439	0.440	0.441	0.443	0.441
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1						

Tabella 43: regressione lineare relativa al tempo di visita

Relativamente alla durata delle visite al contrario, si conferma in maggior parte tutto ciò riscontrato in precedenza, fatta eccezione per la significatività della variabile relativa al genere. In particolare:

- Dalla prima riga non emergono differenze significative a livello di categorie aggregate di siti Media e Social;
- La quarta riga rappresenta come il comportamento degli uomini risulti variato dal regolamento;
- La quinta riga indica omogeneità di comportamento tra le fasce d'età.



VARIABLES	(1) log_pages_ visited	(2) log_pages_ visited	(3) log_pages_ visited	(4) log_pages_ visited	(5) log_pages_ visited	(6) log_pages_ visited
postxtreat	-0.0329 (0.0417)	-0.0613** (0.0300)	-0.0323 (0.0279)	-0.0322 (0.0281)	-0.100 (0.161)	-0.00178 (0.0790)
treat	0.107* (0.0552)	0.0723 (0.0533)	0.0538 (0.0525)			
post	0.00986 (0.0390)	0.0188 (0.0268)				
genderxpostxtreat					0.0719 (0.157)	
agexpostxtreat						-0.0321 (0.0803)
Constant	16.92*** (0.0497)	16.95*** (0.0374)	16.98*** (0.0535)	17.07*** (0.0756)	17.08*** (0.0712)	17.07*** (0.0756)
Observations	38,096	38,096	38,096	38,096	38,096	38,096
R-squared	0.002	0.221	0.223	0.223	0.223	0.223
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1						

Tabella 44: regressione lineare relativa al numero di pagine visitate

Relativamente al numero di pagine visitate si può notare come in nessuna delle variabili di studio emergano coefficienti statisticamente significativi, confermando così i risultati ottenuti nei confronti temporali svolti nel capitolo precedente tramite la piattaforma Excel. In particolare:

- Dalla prima riga non emergono differenze significative a livello di categorie aggregate di siti Media e Social;
- La quarta riga rappresenta che non vi sono differenze di comportamenti tra utenti uomini e donne;
- La quinta riga indica omogeneità di comportamento tra le fasce d'età.



VARIABLES	(1) bounce_rate	(2) bounce_rate	(3) bounce_rate	(4) bounce_rate	(5) bounce_rate	(6) bounce_rate
postxtreat	0.00247 (0.0125)	0.00538 (0.00695)	3.16e-05 (0.00645)	-0.000145 (0.00644)	0.0394 (0.0568)	-0.00136 (0.0105)
treat	-0.0372** (0.0170)	-0.0228* (0.0131)	-0.0194 (0.0138)			
post	0.000897 (0.0109)	-0.00120 (0.00643)				
genderxpostxtreat					-0.0420 (0.0573)	
agexpostxtreat						0.00129 (0.0114)
Constant	0.543*** (0.0169)	0.534*** (0.00923)	0.520*** (0.0104)	0.516*** (0.0170)	0.510*** (0.0139)	0.516*** (0.0170)
Observations	38,096	38,096	38,096	38,096	38,096	38,096
R-squared	0.010	0.888	0.889	0.890	0.890	0.890
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1						

Tabella 45: regressione lineare relativa al bounce rate

Da quest'ultima tabella si vede come i risultati ricalchino quelli emersi relativamente al traffico: la negazione dei trend temporali e la non significatività di nessuna delle variabili di studio.



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).

In primo luogo, 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 all'analisi dei trend temporali sono emersi i risultati seguenti.

A livello macro si può affermare come il traffico europeo abbia subito un calo netto tra prima e dopo l'emissione, soprattutto se paragonato al traffico statunitense, risultato simile è apparso per quanto riguarda la frequenza di rimbalzo che è cresciuta in maniera rilevante, mentre quella degli USA ha subito una crescita nettamente inferiore. Al contrario, per quanto riguarda le statistiche relative alla durata delle visite e al numero di pagine visitate non si riscontrano differenze sostanziali tra prima e dopo l'emissione del regolamento europeo, i dati risultano infatti mantenere un trend simile sia in Europa sia in America. Una spiegazione a questa differenza tra le quattro statistiche può essere ricondotta al fatto che le prime due rappresentano l'iterazione iniziale con il sito, cioè l'arrivo sul dominio (traffico) e l'iterazione con esso (Bounce Rate), quindi vengono influenzati il primo dalle notizie in tema GDPR, il secondo soprattutto dai banner obbligatori che i siti devono esporre sulla loro "landing page",



al contrario gli altri due indicatori rappresentano le iterazioni successive tra utente e dominio, quindi il loro valore non viene determinato da tutti gli utenti ma soltanto da coloro che restano sul dominio una volta accettate le sue condizioni, e per questo il comportamento di questa tipologia di persone non varia tra prima e dopo l'emissione; Relativamente al confronto tra Media e Social: entrambe le categorie presentano i risultati visti in precedenza con i dati aggregati, cioè una riduzione del traffico accompagnata da un aumento della frequenza di rimbalzo molto più elevati in UE, con valori simili per entrambe le classi, per questo si può affermare che non vi è stata alcuna differenza di impatto, si manifesta omogeneità.

La seconda analisi di dettaglio, relativa alle differenze comportamentali in base al genere di appartenenza della maggioranza degli utenti del sito ha portato, invece, a rilevare un'eterogeneità nei valori relativi al traffico. Per entrambi i sessi questo parametro mostra un'importante riduzione successiva al GDPR, ottenuta anche in questo caso tramite il confronto con gli Stati Uniti, ma risulta molto più alta per il genere femminile. Questo, infatti, mostra una riduzione media del 10% circa a confronto con quella degli uomini che si attesta sul 5%. Al contrario la frequenza di rimbalzo cresce a livello europeo per entrambi i generi con valori comparabili, mentre per gli ultimi due indicatori, durata media e numero di pagine visitate, non si manifestano effetti rilevanti a seguito dell'entrata in vigore del regolamento. Una possibile spiegazione a questa eterogeneità presente si può ipotizzare nella maggiore sensibilità delle donne ai temi riguardanti la loro privacy, conseguenza tra gli altri fattori del maggior numero di attacchi informatici a cui sono sottoposte

L'ultima analisi mette a confronto le varie fasce d'età per valutare eventuali eterogeneità presenti tra esse che portino a evidenziare effetti diversi avuti dal GDPR. Si possono distinguere tre diversi comportamenti a seconda dell'età, ottenuti anche in questo caso tramite un confronto tra domini europei e americani:

- I più giovani (sotto i 24 anni) non manifestano differenze significative dovute al regolamento europeo in nessuno degli indicatori presenti;
- Gli utenti compresi tra i 25 e i 65 anni, al contrario, mostrano una forte riduzione del traffico e un costante aumento della frequenza di rimbalzo;
- Sopra i 65 anni si evidenzia soltanto un aumento significativo della frequenza di rimbalzo.



**Politecnico
di Torino**

Riassumendo si può affermare come sicuramente gli utenti compresi nella fascia cosiddetta “lavorativa” (25-65 anni) abbiano una sensibilità ai temi analizzati molto elevata, mentre i più giovani e i più anziani siano poco toccati da argomenti relativi alla privacy online. Le motivazioni possono essere varie, ma sicuramente una di esse deriva dalla tipologia di dati in possesso di coloro che svolgono una professione, questo tipo di dati, infatti, non sono dati personali ma dati appunto lavorativi e per questo soggetti a un’attenzione sicuramente più alta. Un secondo fattore può essere il generale disinteresse manifestato dai giovani per queste tematiche che li porta ad accettare senza leggere la maggior parte dei banner presenti sui siti. Ad ultimo la differenza tra i valori di traffico e frequenza di rimbalzo per i più anziani trova motivazione nella bassa informazione e conoscenza che questo tipo di utenti possiede in ambito privacy, che li porta a non variare il numero di visite ai domini, e al fatto che sono però portati a leggere attentamente i disclaimer obbligatori che i siti devono presentare e spesso dopo averlo fatto decidono di terminare la loro iterazione perché “spaventati” da questi.

Infine, riguardo ai modelli di regressione lineare non si sono evidenziate statistiche significative per nessuna delle ricerche effettuate, fatta eccezione per la durata delle visite che risulta variata statisticamente per gli utenti di genere maschile.

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.



Appendice

Appendice 1

Di seguito un elenco dei siti di News e Media individuati:

163.com	detroitnews.com	kataweb.it	nicolaporro.it	stern.de
1limburg.nl	digg.com	kentonline.co.uk	nieuws.nl	stltoday.com
20minutes.fr	dilei.it	krone.at	nj.com	stol.it
abc7.com	dna.fr	ksl.com	nola.com	stuttgarter-nachrichten.de
abc7news.com	donnamoderna.com	ksta.de	noordhollandsdagblad.nl	supereva.it
abendblatt.de	drimble.nl	la-croix.com	nordbayern.de	tag24.de
abendzeitung-muenchen.de	drudgereport.com	ladbible.com	notizie.it	tagesschau.de
activehosted.com	dvhn.nl	ladepeche.fr	notretemps.com	tagesspiegel.de
actu.fr	dw.com	laleggepertutti.it	nouvelobs.com	taz.de
ad.nl	echo.msk.ru	lamontagne.fr	noz.de	techbook.de
adnkronos.com	ecodibergamo.it	lanazione.it	nrc.nl	telegraaf.nl
affaritaliani.it	economist.com	lankasri.com	n-tv.de	telegraph.co.uk
agi.it	ed.nl	lanouvellerepublique.fr	nu.nl	tfl.gov.uk
ajc.com	eenadu.net	larena.it	nw.de	theatlantic.com
al.com	egaliteetreconciliations.fr	lastampa.it	nydailynews.com	theblaze.com
aljazeera.com	elle.fr	latimes.com	nymag.com	thecut.com
allestoringen.nl	esquire.com	latribune.fr	nypost.com	thedailybeast.com
alot.com	estrepubicain.fr	lc.nl	nytimes.com	theepochtimes.com
americanthinker.com	europe1.fr	ledauphine.com	nzz.ch	thegatewaypundit.com
ansa.it	express.co.uk	lefigaro.fr	o2.pl	theguardian.com
apnews.com	express.de	leggo.it	olivesoftware.com	thehill.com
apotheken-umschau.de	fanpage.it	lemonde.fr	onet.pl	thesun.co.uk
augsburger-allgemeine.de	fark.com	lenta.ru	onlyinyourstate.com	thetimes.co.uk
aujourd'hui.fr	fastcompany.com	leparisien.fr	oozo.nl	theverge.com
avrotros.nl	faz.net	lepoint.fr	order-order.com	theweek.com
avvenire.it	femmeactuelle.fr	leprogres.fr	oregonlive.com	thewirecutter.com
axios.com	focus.de	lesechos.fr	oreilly.com	theyeshivaworld.com
azcentral.com	focus.it	liberation.fr	osthessen-news.de	thisismoney.co.uk
badische-zeitung.de	fok.nl	libero.it	ouest-france.fr	thomsonreuters.com



**Politecnico
di Torino**

bbc.co.uk	forbes.com	liberoquotidiano.it	palermotoday.it	thrillist.com
bbc.com	foxnews.com	limburger.nl	panorama.it	time.com
bd.nl	freep.com	lindependant.fr	parismatch.com	tiscali.it
belfasttelegraph.co.uk	ft.com	linkiesta.it	parool.nl	today.com
berliner-zeitung.de	gazeta.pl	linternaute.com	patch.com	today.it
bienpublic.com	gazzettadiparma.it	liverpoolecho.co.uk	people.com	t-online.de
bigodino.it	gds.it	livesicilia.it	persgroep.net	topics.nl
bild.de	gelderlander.nl	liveuamap.com	pi-news.net	townhall.com
birminghammail.co.uk	gelocal.it	loopings.nl	pjmedia.com	tpi.it
bizjournals.com	gentside.com	lunion.fr	planet.fr	tpo.nl
blastingnews.com	giornaledibrescia.it	lvz.de	plusonline.nl	trouw.nl
bloomberg.com	globo.com	madmoizelle.com	pnp.de	twentytwowords.com
bndestem.nl	gq.com	makeuseof.com	politico.com	tz.de
bnnvara.nl	greatandhra.com	manchestereveningnews.co.uk	presseportal.de	ultrasurfing.com
boston.com	handelsblatt.com	marianne.net	pzc.nl	unionesarda.it
bostonglobe.com	haz.de	marieclaire.fr	qq.com	usatoday.com
breitbart.com	hessenschau.de	maville.com	quotenet.nl	usinenouvelle.com
buienradar.nl	hln.be	mediapart.fr	quotidiano.net	usnews.com
bunte.de	hna.de	mensenlinq.nl	qz.com	utopia.de
businessinsider.com	hotair.com	mentalfloss.com	rainews.it	valeursactuelles.com
businessonline.it	huaren.us	mercurynews.com	rambler.ru	vanityfair.com
bustle.com	huffingtonpost.co.uk	merkur.de	rawstory.com	vanityfair.it
buzzfeed.com	huffpost.com	mesopinions.com	rbb24.de	varesenews.it
cbslocal.com	icepop.com	meteolive.it	rd.com	variety.com
cbsnews.com	ighome.com	metro.co.uk	realclearpolitics.com	varmatin.com
centurylink.net	ilfattoquotidiano.it	metronieuws.nl	realsimple.com	vice.com
charentelibre.fr	ilfoglio.it	mewe.com	refinery29.com	virgilio.it
chicagotribune.com	ilgazzettino.it	miamiherald.com	repubblica.it	vnexpress.net
chosun.com	ilgiornale.it	milanotoday.it	reuters.com	volkskrant.nl
chron.com	ilgiornaledivicenza.it	mirror.co.uk	romatoday.it	vox.com
chroniclelive.co.uk	ilgiorno.it	mitbbs.com	rp-online.de	waldnet.nl
cleveland.com	ilmattino.it	mlive.com	rt.com	walesonline.co.uk
cnn.com	ilmessaggero.it	mopo.de	rtlnews.nl	want.nl
complex.com	ilrestodelcarlino.it	msn.com	sat24.com	washingtonpost.com
corriere.it	ilsecoloxix.it	msnbc.com	sciencesetavenir.fr	washingtontimes.com
	ilsussidiario.net	mumsnet.com	scmp.com	waz.de



cosmopolitan.com	iltempo.it	mypass.de	scotsman.com	wenxuecity.com
cosmopolitan.fr	indebuurt.nl	mz-web.de	seattletimes.com	westernjournal.com
dailycaller.com	indiatimes.com	nationalreview.com	sfgate.com	worldjournal.com
dailydot.com	inews.co.uk	naver.com	shz.de	wp.pl
dailykos.com	inquirer.com	nbcnews.com	sina.com.cn	wral.com
dailymail.co.uk	interia.pl	ndtv.com	slate.fr	wsj.com
dailyrecord.co.uk	internazionale.it	newsela.com	smithsonianmag.com	wunderweib.de
dailystar.co.uk	iodonna.it	newsmax.com	sohu.com	yahoo.co.jp
dailywire.com	irishtimes.com	newsnow.co.uk	southernliving.com	yahoo.com
dallasnews.com	jeugdjournaal.nl	newspapers.com	spectator.co.uk	ynet.co.il
decorrespondent.nl	journalistenwatch.com	newsweek.com	sputniknews.com	zeit.de
denverpost.com	jungefreiheit.de	newyorker.com	stadt-bremerhaven.de	zerohedge.com
der-postillon.com	kachelmannwetter.com	nhnieuws.nl	standard.co.uk	watson.de
destentor.nl	kalender-365.nl	nicematin.com	startribune.com	

Tabella 46: elenco siti News&Media

Mentre i Social Network:

bp.blogspot.com	foursquare.com	messenger.com	proboards.com	tumblr.com
classmates.com	icv-crew.org	nebenan.de	redd.it	twitter.com
computerbase.de	instagram.com	nextdoor.com	reddit.com	umbl.com
developpez.net	jappy.com	ok.ru	skyrock.com	viadeo.com
drwindows.de	linkedin.com	omgle.com	slack.com	vk.com
facebook.com	livejournal.com	pinterest.co.uk	slideshare.net	weibo.com
fiverr.com	lovoo.com	pinterest.com	spiceworks.com	whatsapp.com
forumactif.com	medium.com	pinterest.de	spin.de	yammer.com
forumconstruire.com	meetup.com	pinterest.fr	tagged.com	zhihu.com

Tabella 47: elenco siti Social Network



Appendice 2

Vengono di seguito riportati per ogni categoria esaminata nel paragrafo 4.3 i grafici non rilevanti ai fini dell'analisi in quanto non presentanti variazioni significative:

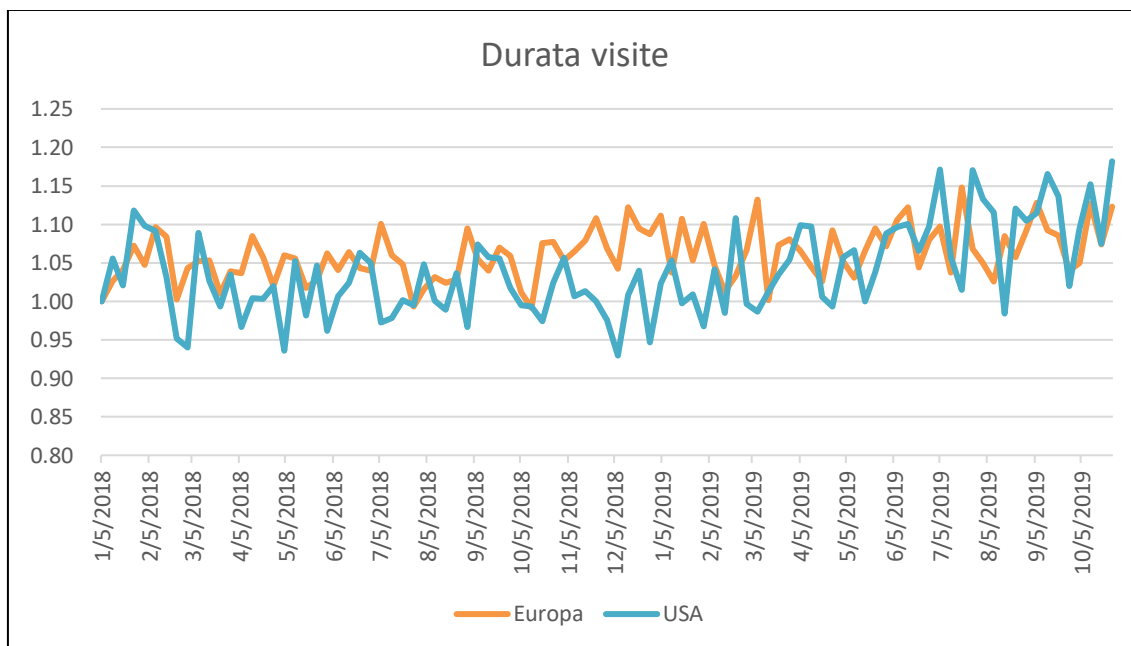


Tabella 48: Confronto durata visite News&Media tra Europa e USA

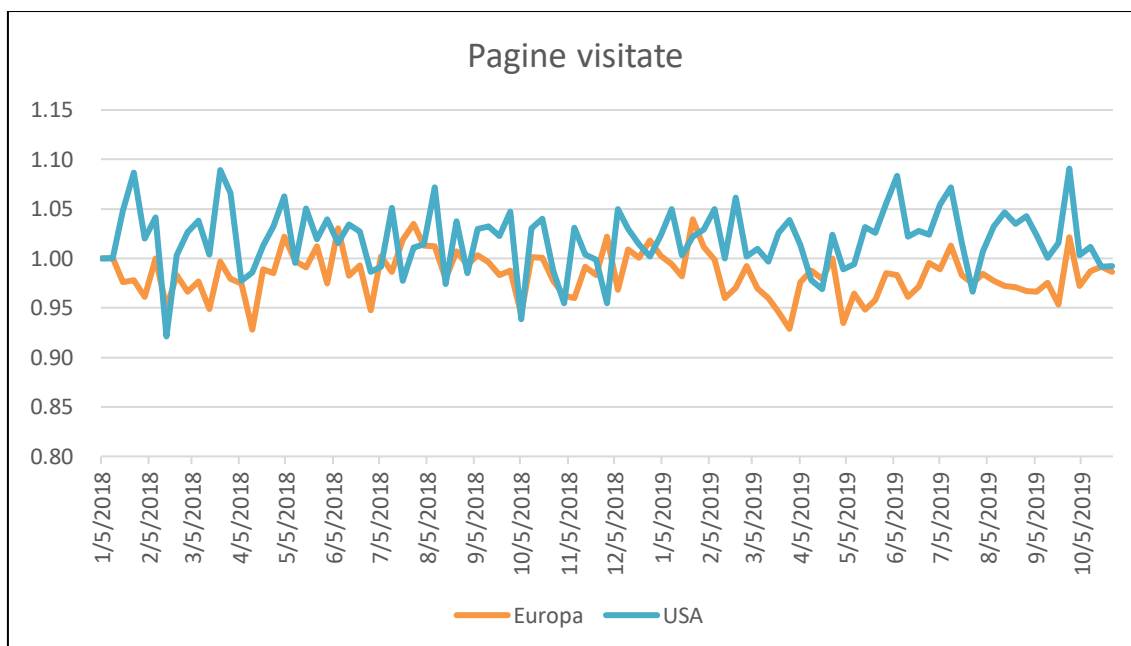


Tabella 49: Confronto pagine visitate News&Media tra Europa e USA

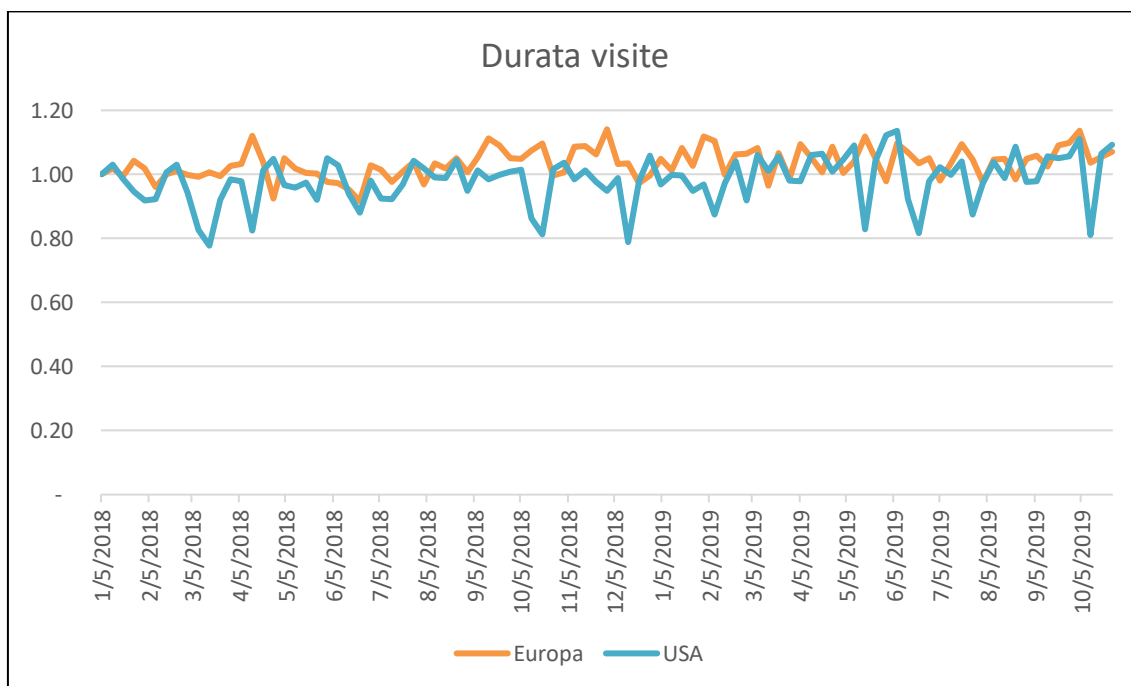


Tabella 50: Confronto durata visite Social Network tra Europa e USA

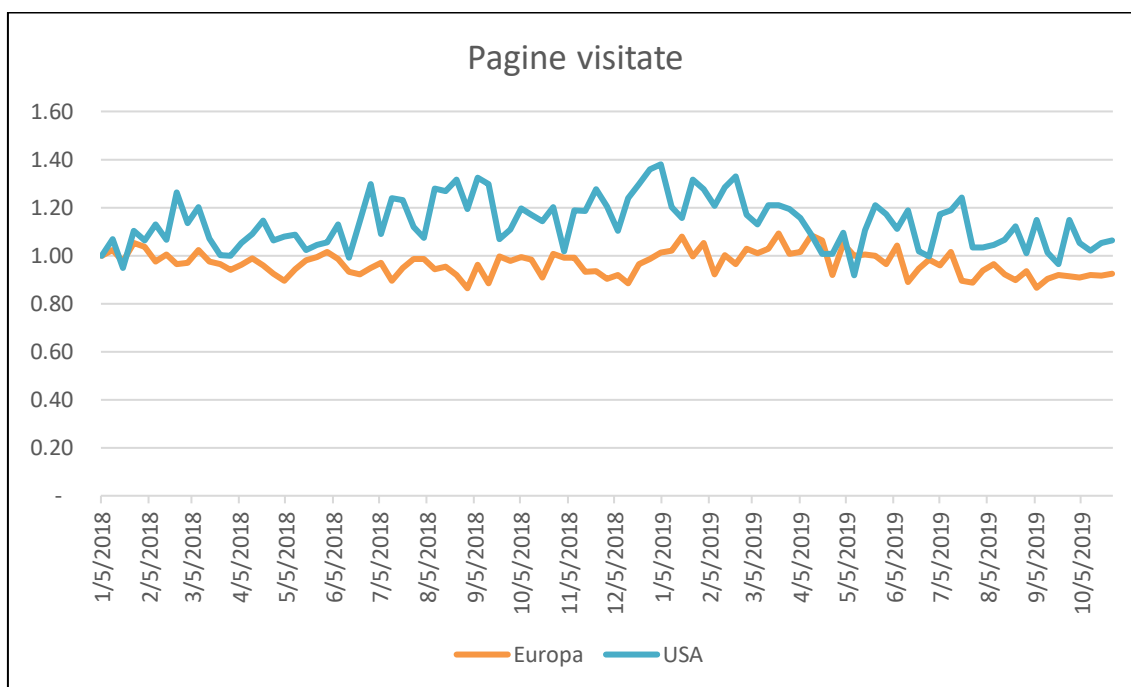


Tabella 51: Confronto pagine visitate Social Network tra Europa e USA

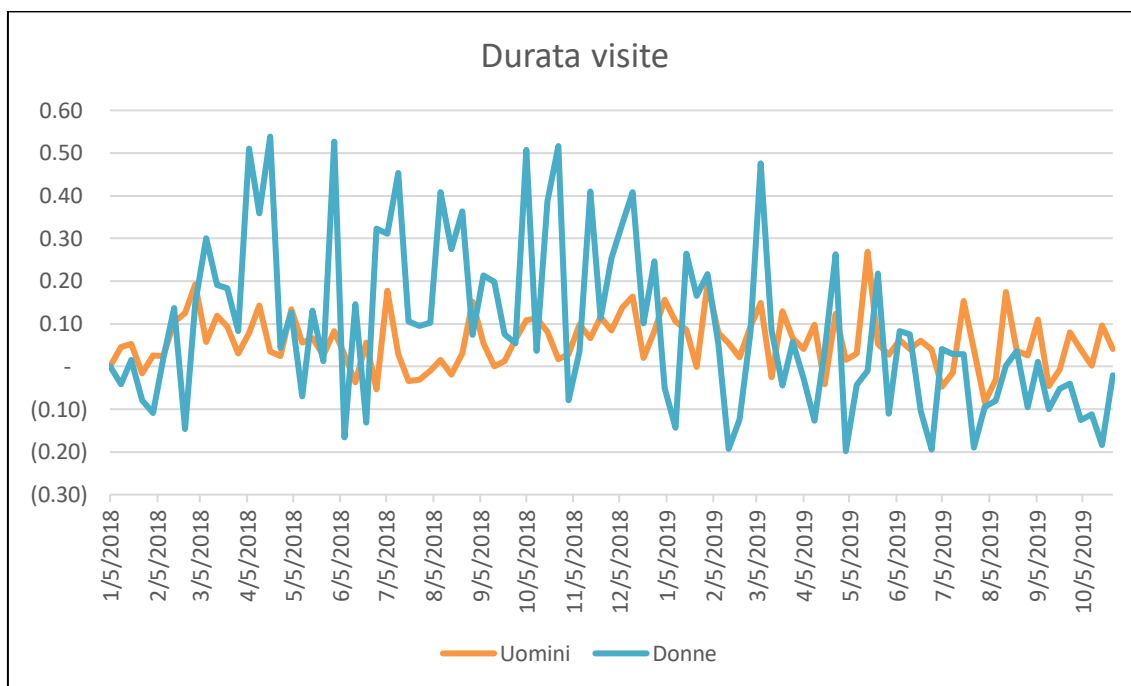


Tabella 52: Confronto differenziale durata visite tra uomini e donne europei

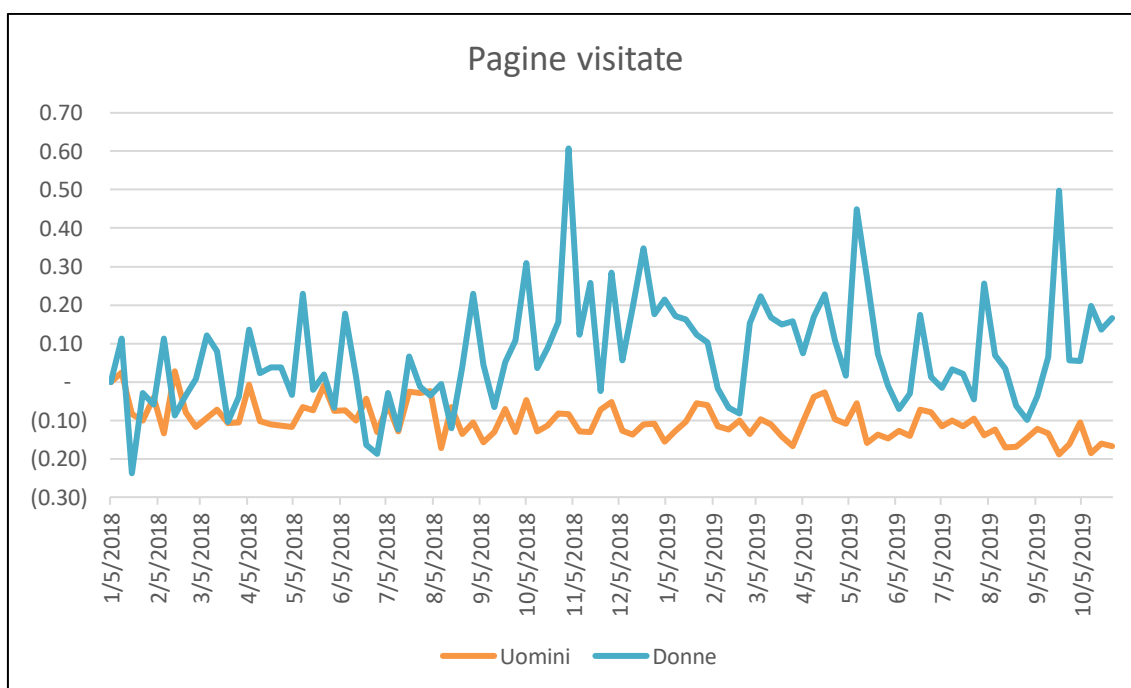


Tabella 53: Confronto differenziale pagine visitate tra uomini e donne europei

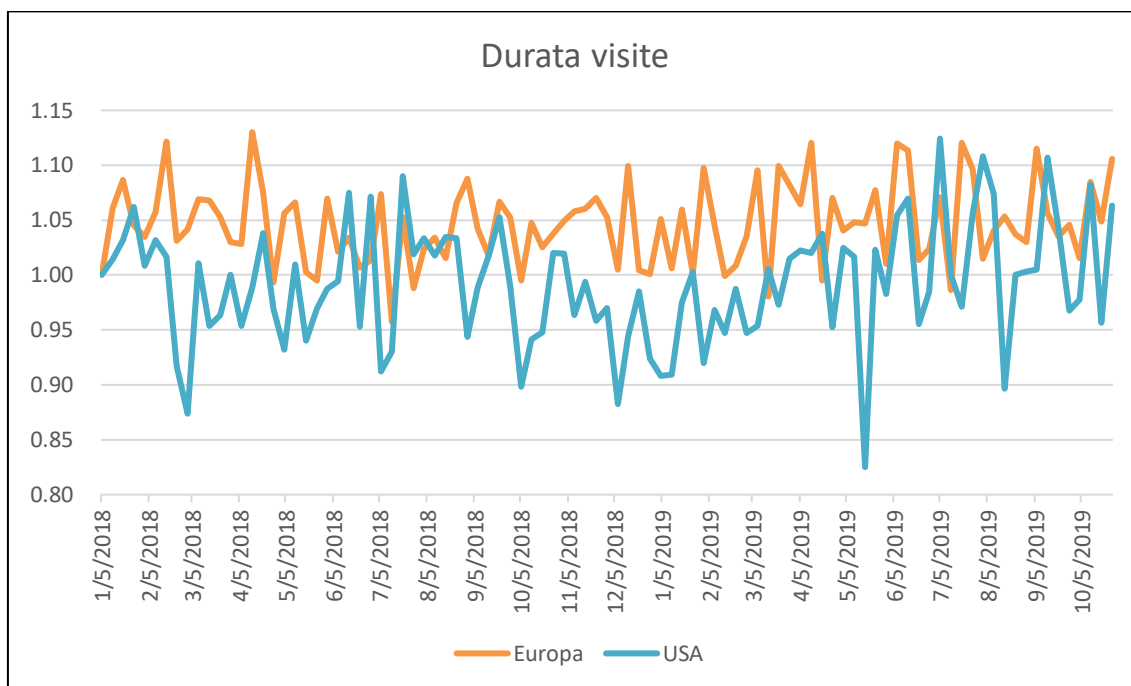


Tabella 54: Confronto durata visite uomini tra Europa e USA

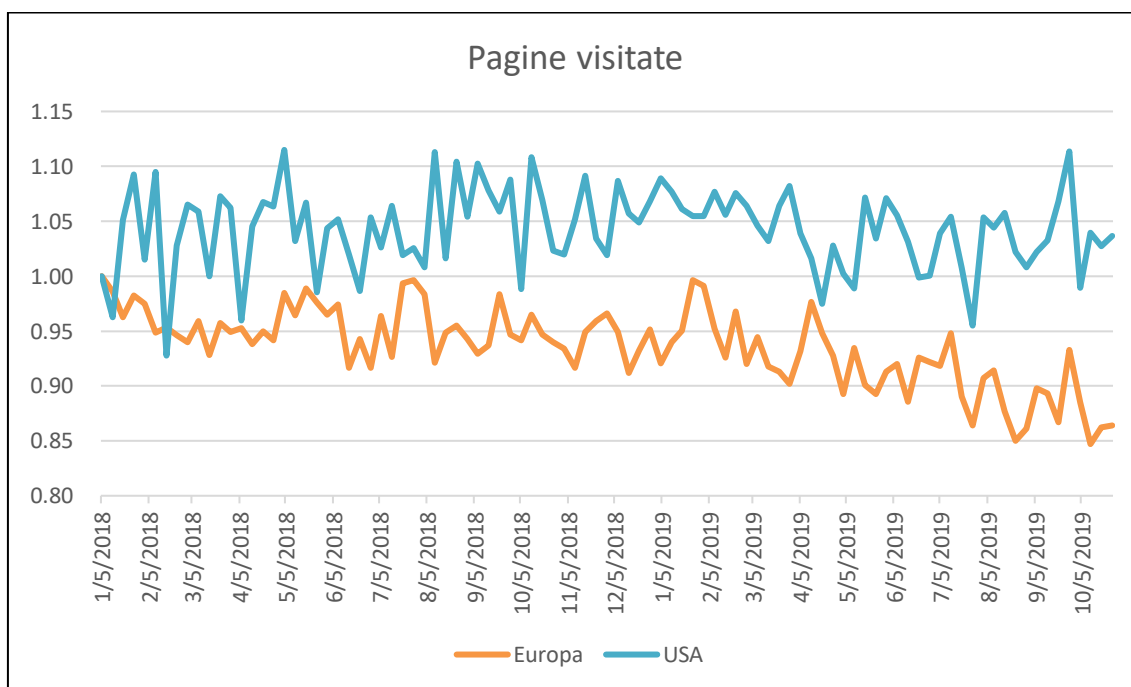


Tabella 55: Confronto pagine visitate uomini tra Europa e USA

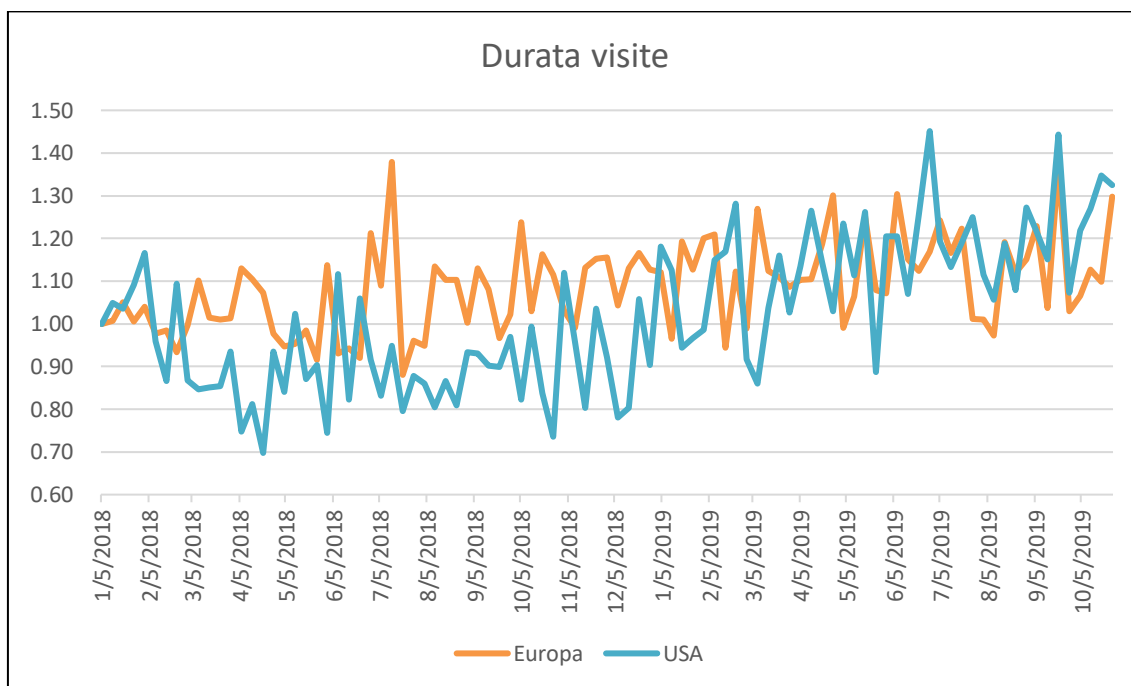


Tabella 56: Confronto durata visite donne tra Europa e USA

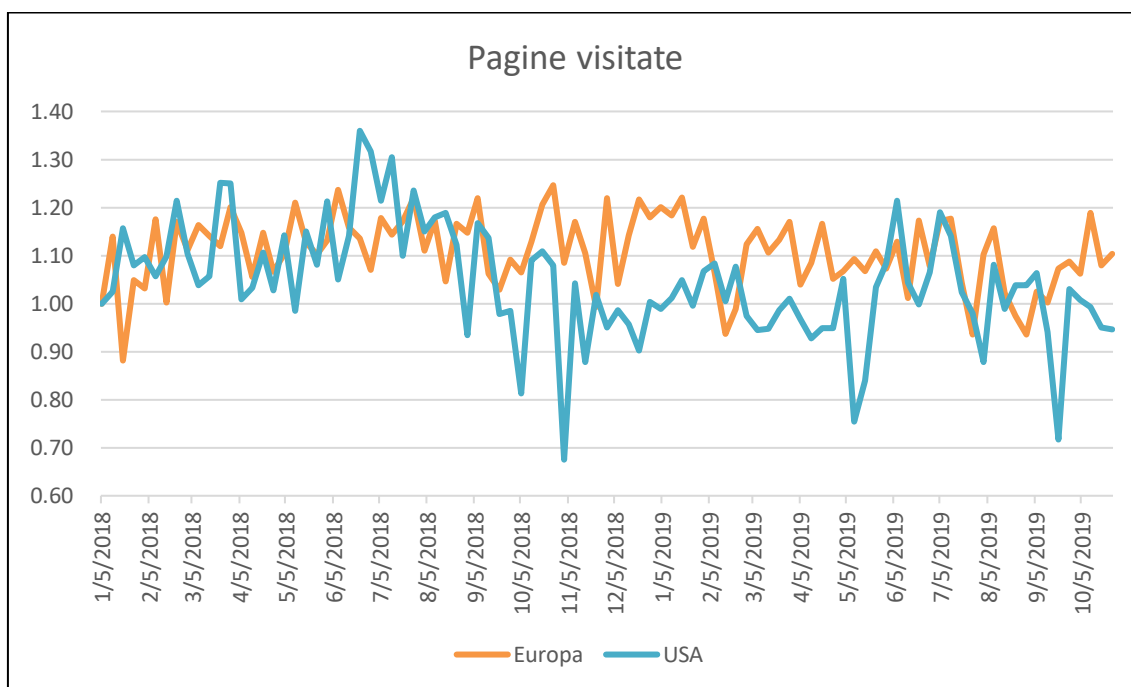


Tabella 57: Confronto pagine visitate donne tra Europa e USA



Politecnico
di Torino

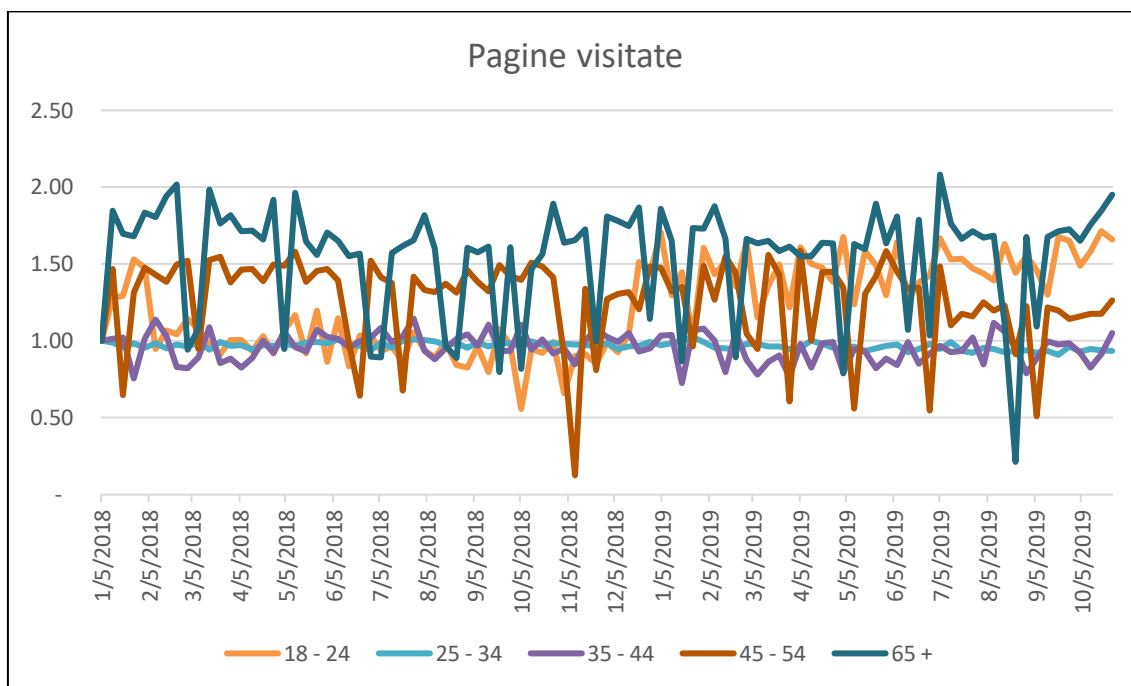


Tabella 58: confronto pagine visitate tra fasce d'età a livello europeo

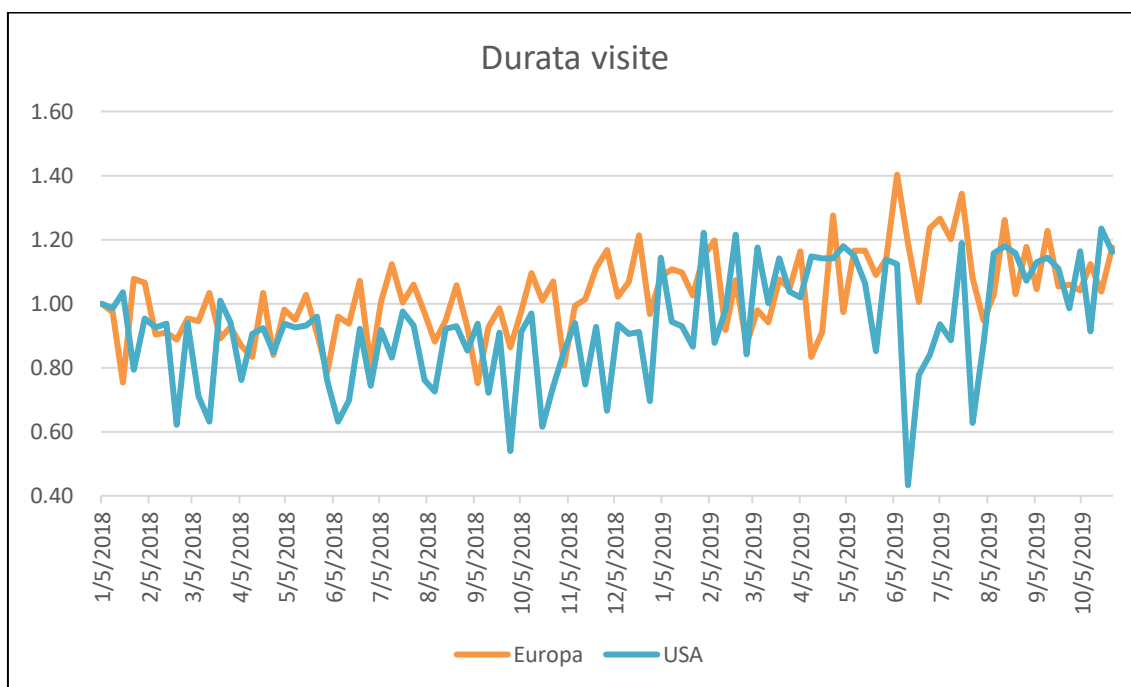


Tabella 59: confronto durata visite 18-24 anni tra Europa e USA

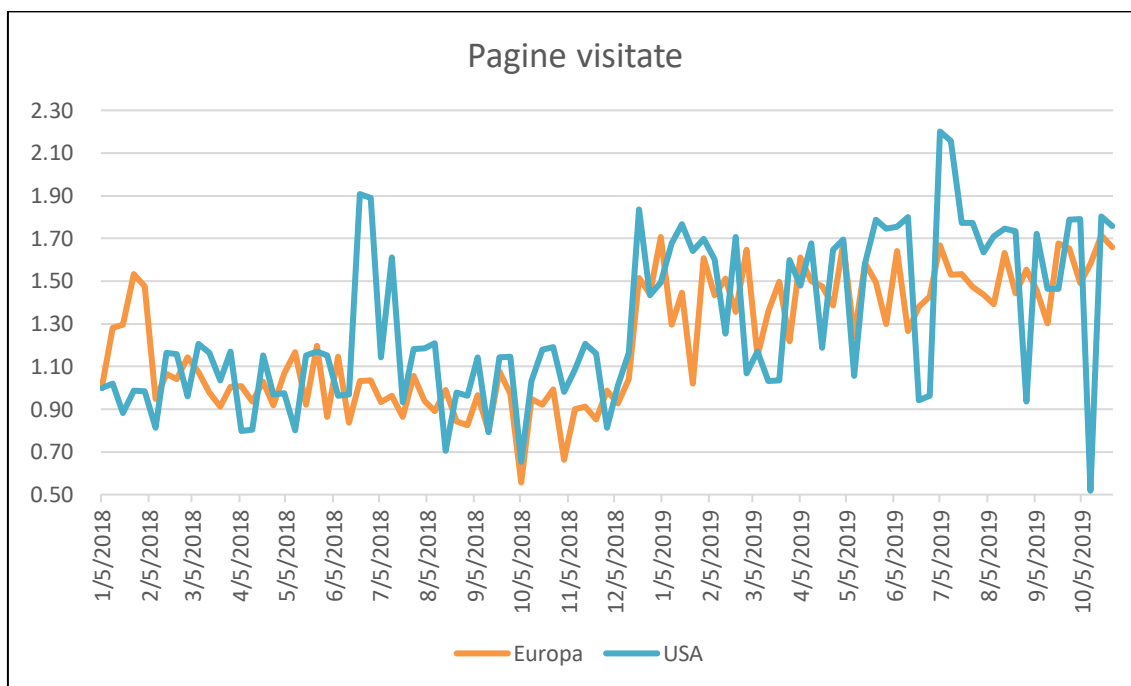


Tabella 60: confronto durata visite 18-24 anni tra Europa e USA

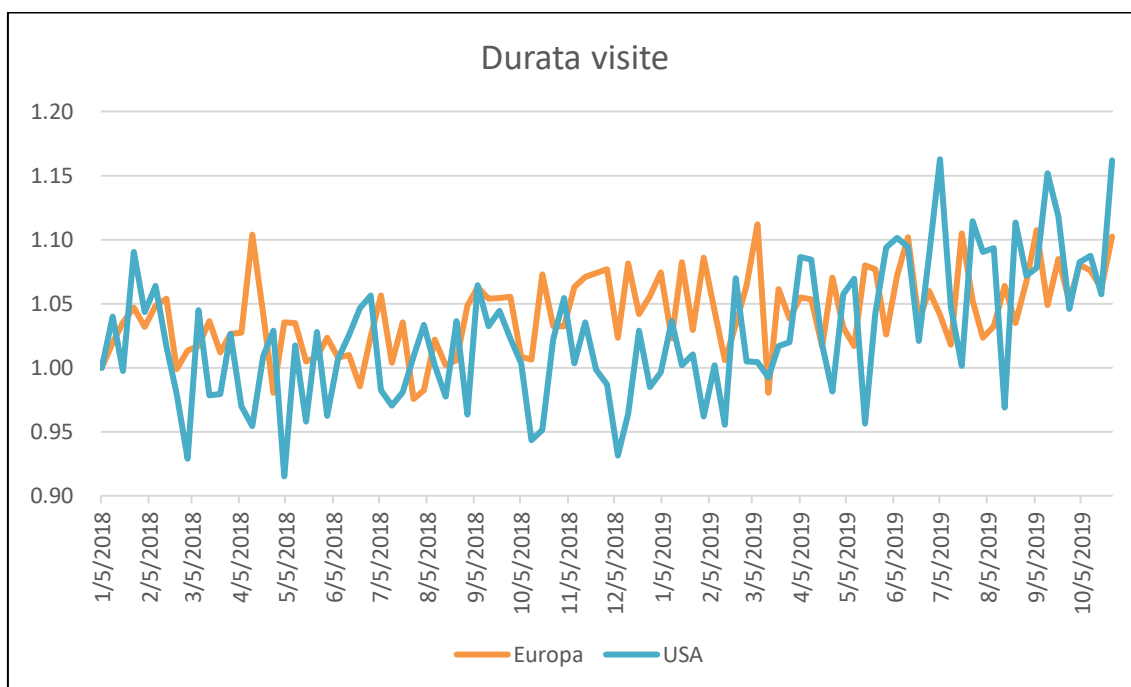


Tabella 61: confronto durata visite 25-34 anni tra Europa e USA

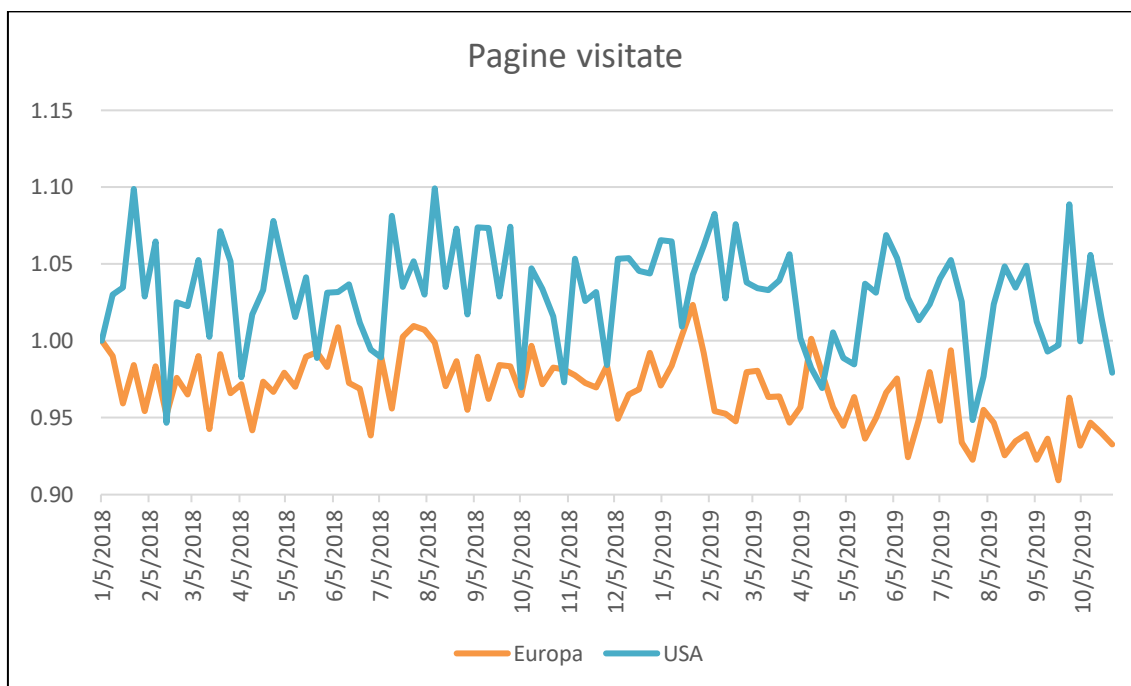


Tabella 62: confronto pagine visitate 25-34 anni tra Europa e USA

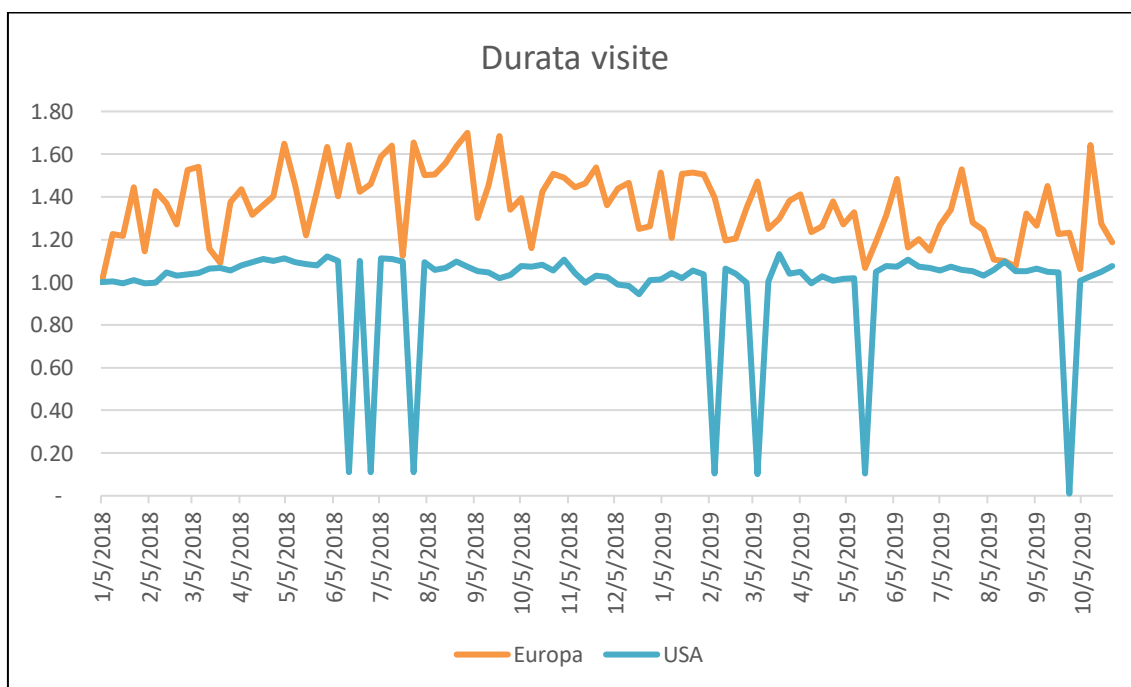


Tabella 63: confronto durata visite 35-44 anni tra Europa e USA



Politecnico
di Torino

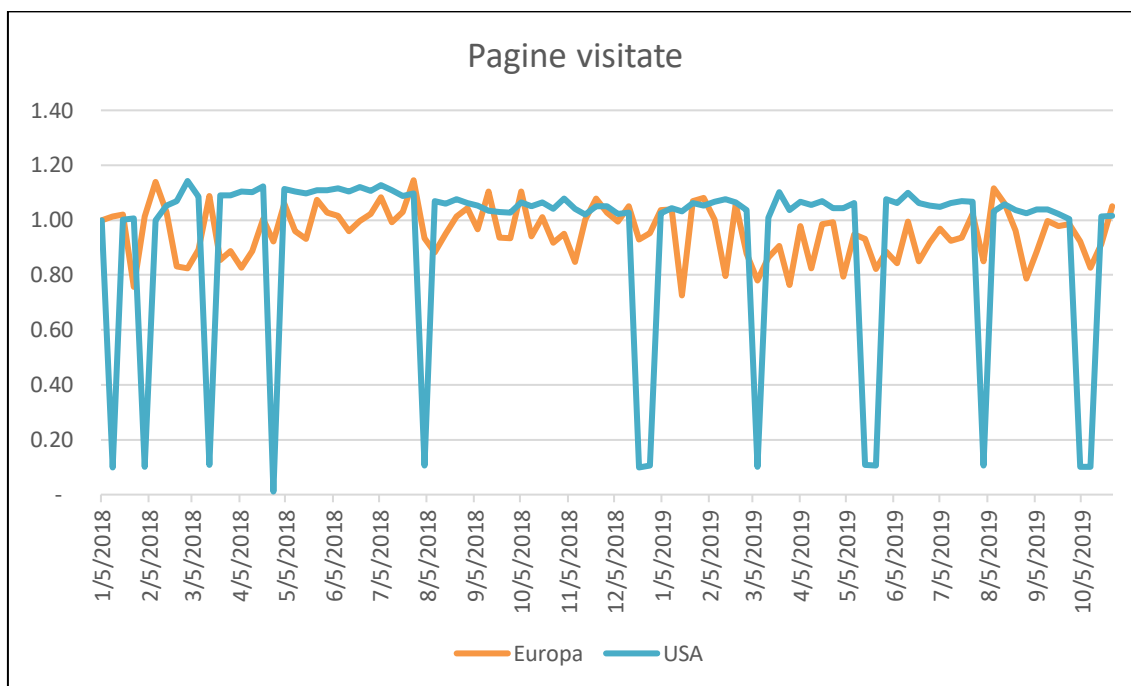


Tabella 64: confronto pagine visitate 35-44 anni tra Europa e USA

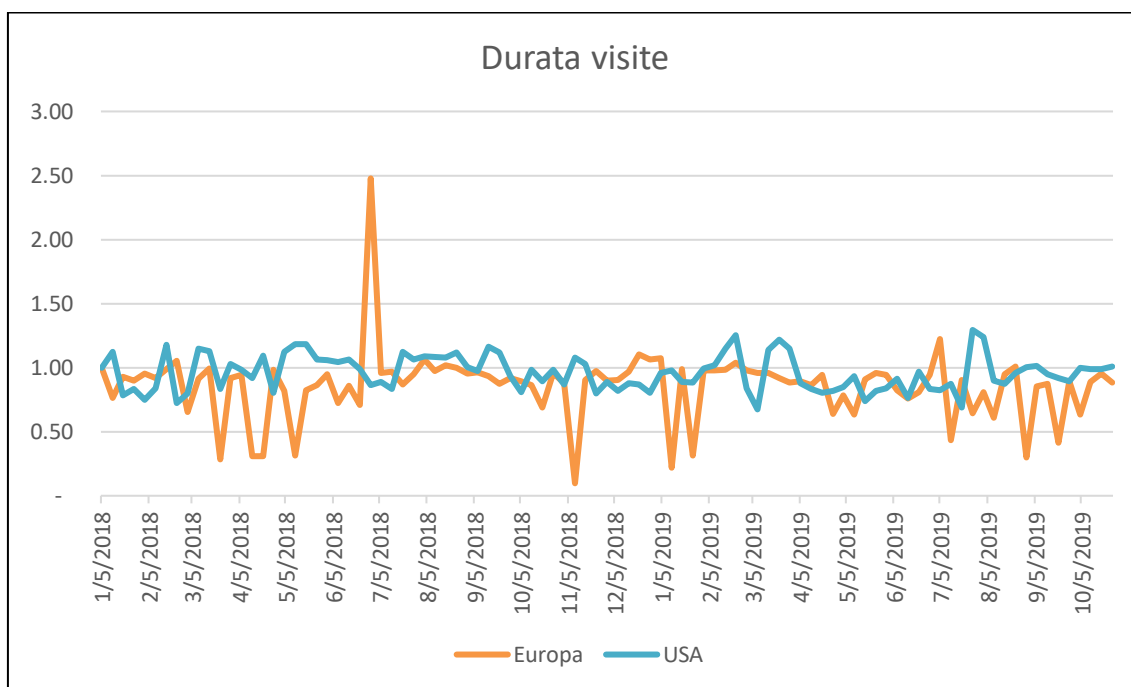


Tabella 65: confronto durata visite 65+ anni tra Europa e USA



Politecnico
di Torino

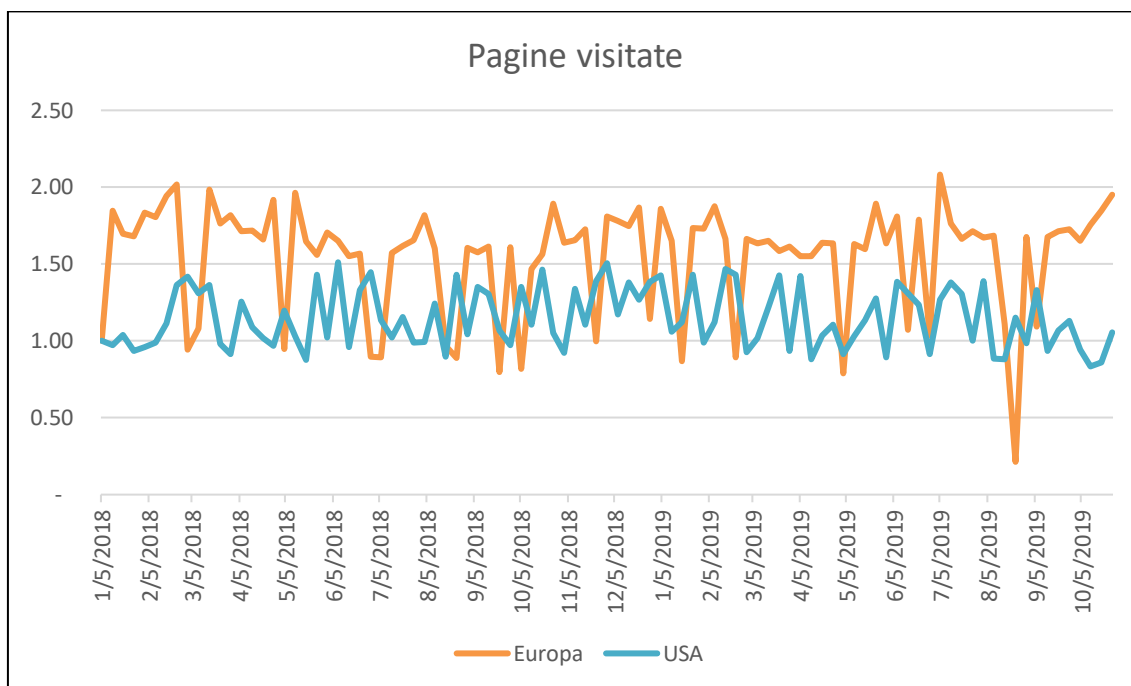


Tabella 66: confronto pagine visitate 65+ anni tra Europa e USA



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. and 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. 2020.** *Protecting Citizens' Personal Data and privacy: Joint Effort from GDPR EU Cluster Research Projects.*
- Ding, X., et al. 2015.** *A fair data market system with data quality evaluation and repairing recommendation.* 2015.
- Fast, V., Schnurr, D. and 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 research 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. and Sapi, G. 2021.** *Online Privacy and Market Structure: Theory and Evidence.* 2021.
- Sabatino, L., Congiu, R. and 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. and Nussbaum. 2020.** *Laws versus technology: Blockchain, GDPR and tough tradeoffs.* 2020.



**Politecnico
di Torino**

Thai, M. T., Wu, W. and Xiong, H. 2016. *Big Data in complex and social networks.* 2016.

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