UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DIPARTIMENTO DI ECONOMIA E MANAGEMENT

Corso di Laurea
in ECONOMIA E AZIENDA DIGITALE

Relazione Finale

L'esperienza dei visitatori museali: Un'indagine sui post Instagram e sulle recensioni

Relatore: Chiar.mo Prof. Enrico Ripamonti

Laureando:
Mattia Natali

Matricola n. 719573

Anno Accademico 2021/2022

Indice

Introduzione 1

1 Il settore museale 3

1.1 I musei nel mondo 3

1.1.1 Il ruolo sociale ed economico dei musei 4

1.2 I musei in Italia 7

1.3 Musei e covid 8

1.4 Musei e social media 9

1.5 Musei e analisi dati 12

1.5.1 Analisi spostamenti visitatori nel Louvre 14

1.5.2 Analisi reazioni emotive Bologna Musei 14

2 Indagine sul comportamento dei visitatori analizzando i post di Instagram 15

2.1 Introduzione 15

2.2 Ricerche precedenti 16

2.3 Instagram 17

2.4 Dati considerati 17

2.5 Analisi hashtag 18

2.6 Analisi immagini 20

2.7 Sintesi analisi delle immagini 30

3 Analisi recensioni TripAdvisor 32

3.1 Perché analizzare le recensioni 32

3.2 TripAdvisor 32

3.3 Importanza delle recensioni 33

3.3.1 Recensioni false 34

3.4 Analisi delle recensioni 35
3.4.1 Analisi delle parole 36
3.4.2 Sentiment analysis 39
3.4.3 Analisi con bigrammi 43
3.4.4 Analisi tf-idf 44
Conclusioni 46
Bibliografia 50
Sitografia 54
Indice delle figure 56
Indice delle tabelle 58

Introduzione

L'importanza della cultura, dell'arte, della storia e della scienza è indubbia e i musei sono i luoghi cardine in cui esse sono tramandate. Sono istituzioni estremamente importanti e fondamentali sia per motivi culturali che sociali, ma anche centri d'interesse economico in quanto elementi di attrazione turistica.

Essendo spesso istituzioni senza fini di lucro essi fanno affidamento su donazioni e i fondi conferiti dai governi. Nei confronti di questi ultimi hanno il compito di dimostrare di valere il costo sostenuto dalla collettività. Per questo i musei sono sempre alla ricerca di metodi per comprendere il comportamento e l'interesse dei turisti in modo da migliorare la propria offerta e allargare il proprio bacino d'utenza. Le nuove tecnologie sembrano essere una nuova strada percorribile, non solo per fornire una nuova tipologia di esperienza, ma anche per ottenere importanti informazioni sui visitatori. Vi è però un problema sui costi che questi strumenti comportano, cosa che molti musei non possono permettersi. Sicuramente, però, la presenza sui social media è alla portata di tutti ed è difficile pensare ad una galleria che non ne abbia. Queste piattaforme offrono strumenti integrati molto utili per conoscere il proprio pubblico, i quali possono essere integrati con sistemi e software esterni per estrapolare ulteriori informazioni, facilmente recuperabili.

La tesi in questione si articola in tre capitoli e una conclusione finale sui risultati ottenuti.

Il primo capitolo ha come obiettivo di introdurre il settore museale, il suo ruolo economico, sociale e le nuove sfide portate dalla pandemia globale di Covid19, per poi porre l'attenzione su come utilizzare i social media e nuove tecnologie per raccogliere informazioni sui visitatori e il loro comportamento per riuscire poi a sviluppare strategie per migliorare l'offerta e attrarre nuovi turisti.

Nel secondo capitolo viene proposto un possibile strumento per la raccolta di dati sfruttando i post di Instagram. In questa sezione viene riproposta e parzialmente rielaborata la ricerca "*Analyzing the museum experience through the lens of Instagram posts*" di Rhee, B. A., Pianzola, F., and Choi, G. T. del 2020. In questo report vengono proposte due analisi: la prima osservando le informazioni ricavabili dagli hashtag e la

seconda mediante l'utilizzo di Google Vision API per riconoscere ciò che viene ritratto nelle foto. Per finire vengono create delle categorie per osservare quali sono le aree di interesse che vengono maggiormente fotografate dagli utenti.

Nel terzo capitolo si procede analizzando le recensioni recuperabili sul sito di TripAdvisor riguardanti il MoMA. Viene proposta un'analisi sulle parole utilizzate e una *sentiment analysis* per riuscire a ricavare indicazioni circa le opinioni, le critiche, i punti di forza e la risposta emotiva di coloro che hanno visitato il museo.

Infine, la tesi si conclude con alcune considerazioni finali.

1 Il settore museale

1.1 I musei nel mondo

I musei sono per eccellenza la dimora della cultura, essenziali per tramandarla di generazione in generazione, salvaguardando ed esponendo artefatti di importanza artistica, storica, scientifica e culturale.

ICOM (International Council Of Museum) li definisce in questo modo: "*Il museo è un'istituzione permanente senza scopo di lucro e al servizio della società, che effettua ricerche, colleziona, conserva, interpreta ed espone il patrimonio materiale e immateriale. Aperti al pubblico, accessibili e inclusivi, i musei promuovono la diversità e la sostenibilità. Operano e comunicano eticamente e professionalmente e con la partecipazione delle comunità, offrendo esperienze diversificate per l'educazione, il piacere, la riflessione e la condivisione di conoscenze*"[1].

In essa sono contenute le funzioni fondamentali di un museo: ricerca, collezione, conservazione, interpretazione, comunicazione ed esposizione.

Nel mondo ci sono circa 104 mila musei, di cui il 61% nell'Europa occidentale e in Nord America (UNESCO, 2021).

Nel 2021 il numero di visitatori presso i primi cento musei al mondo è stato di 71 milioni contro i 54 milioni dell'anno precedente, indicando un segnale di ripresa, ma che resta lontano dai 230 milioni del 2019. I musei più visitati, stando al numero di frequentatori dichiarati dalle stesse istituzioni nel 2021 sono stati:

- Louvre, Francia
- Museo Nazionale della Cina, Cina
- Museo Russo, Russia
- Multimedia Art Museum, Russia
- Metropolitan Museum of Art, Stati Uniti
- Galleria degli Uffizi, Italia
- National Gallery of Art, Stati Uniti

[1] https://www.icom-italia.org/definizione-di-museo-di-icom/

- Museo Statale Ermitage, Russia
- Museo Reina Sofía, Spagna
- Musei Vaticani, Città del Vaticano

Sebbene non esista una vera e propria classificazione per i musei, si possono distinguere a seconda della categoria principale delle loro esposizioni, potremmo così differenziarli in:

- arti figurative
- arti applicate
- artigianato
- archeologia
- antropologia ed etnologia
- storia
- storia militare
- scienze
- tecnologia
- museo dei bambini
- storia naturale
- numismatica
- giardini botanici e zoologici
- filatelia

i quali, poi, a loro volta possono specializzarsi in una specifica sottocategoria, per esempio un museo d'arte focalizzato sull'arte moderna.

1.1.1 Il ruolo sociale ed economico dei musei

Nel considerare le funzioni dei musei non si può fare a meno di considerare il ruolo sociale che essi hanno e di come possono influenzare la comunità.

In periodi di contrasti, tensioni sociali e politiche i musei possono farci da monito a ciò che si va in contro mostrandoci gli errori commessi in passato e rendere la "historia

magistra vitae" come affermava Cicerone[2] e, in seguito, Machiavelli[3] e non "magistra di nulla", come sosteneva Montale[4], facendoci cadere sempre, ciclicamente, nei medesimi sbagli.

Bisogna considerare anche il potere dei musei di creare un senso di comunità, ruolo a cui anche i musei più piccoli e locali possono contribuire, mostrando e celebrando un comune senso di appartenenza ad un particolare retaggio culturale. Ovviamente possono anche riunire le persone letteralmente proponendo eventi, dibattiti e incontri.

I musei si occupano anche di istruire le nuove generazioni preparando manifestazioni e programmi ideati per i più piccoli per dar loro un'ulteriore possibilità di apprendere oltre alle normali lezioni scolastiche. Sempre più musei stanno promuovendo forme di edutainment rivolte ad un pubblico più giovane per rendere l'apprendimento un'attività divertente e interessante. Ne è un esempio il Museo Santa Giulia di Brescia[5] che assieme ad Atlantyca Enterteinment e la Fondazione Cariplo, hanno ideato una app con protagonista Geronimo Stilton, un popolare personaggio di libri per ragazzi, in grado di ampliare l'esplorazione del museo con un'esperienza digitale di qualità (Figura 1.1).

---

[2] "*Historia vero testis temporum, lux veritatis, vita memoriae, magistra vitae, nuntia vetustatis*". Cicerone, De Oratore II, 9

[3] "*…cosa alcuna che io abbia più cara o tanto stimi quanto la conoscenza delle azioni degli uomini grandi, appresa con una lunga esperienza degli avvenimenti contemporanei e uno studio assiduo degli avvenimenti antichi".* N. Machiavelli, Il Principe, 1532

[4] "*La storia non è magistra/ di niente che ci riguardi."* E. Montale, La Storia, Satura, 1971

[5] https://www.bresciamusei.com/progetto/geronimo-stilton-brescia-musei-adventures/

Figura 1.1 Esempio interfaccia dell'app Geronimo Stilton Adventures

Pur essendo, per la maggior parte, organizzazioni senza fini di lucro, i musei offrono anche un contributo notevole all'economia.

Stimare o dare valore alla cultura è difficile, se non impossibile, in quanto non è un bene economico sottoposto alla legge della domanda e dell'offerta[6].

Il sostegno alla cultura è legittimato dall'idea del fallimento di mercato e dalla presunta presenza di esternalità economiche e sociali positive. Nel corso degli anni sono stati proposti diversi metodi e modelli per quantificare l'impatto economico dei beni culturali e isolarli al netto del sistema complesso in cui sono inseriti, per quanto complicato sia, si possono facilmente osservare gli effetti diretti sul mercato[7]:

- le retribuzioni del personale
- il valore della spesa di funzionamento e manutenzione

[6] ACTORS Italia, Un modello di simulazione dell'impatto economico dei musei/istituti dell'area POIN: metodo, fonte dati, primi risultati, 2015

[7] ACTORS Italia, *Un modello di simulazione dell'impatto economico dei musei/istituti dell'area POIN: metodo, fonte dati, primi risultati*, 2015

- il valore della spesa dei visitatori

Più complicato è stabilire l'ammontare degli effetti indiretti in quanto sono determinati dalla spesa e dagli effetti moltiplicativi che essa determina. Questa domanda aggiuntiva sollecita l'offerta e la produzione da parte di imprese e altri soggetti economici che fanno riferimento ai flussi di visitatori del Museo. Questi effetti indiretti, che potrebbero essere declinati per tipologia e settore di spesa, generano a loro volta reddito che verrà poi speso sul territorio, con maggiori retribuzioni e spesa da parte delle famiglie[8].

1.2 I musei in Italia

L'Italia è un paese ricco di storia e cultura e non a caso è lo stato con il maggior numero di siti riconosciuti dall'UNESCO, 58 (UNESCO, 2022), e conta una grande quantità di musei e siti archeologici, solo quelli statali sono più di 500 (Direzione Generale Musei, 2022).

In Italia sono presenti alcuni tra i musei più famosi al mondo e nel 2021 la protagonista è stata la Galleria degli Uffizi, che comprende il Giardino di Boboli e Palazzo Pitti, con oltre 1,7 milioni di visitatori, superando il Colosseo che ne ha attratti circa 1,6 milioni, rimanendo, però, pur sempre primo per gli incassi, chiude il podio Pompei con poco più di un milione di turisti[9].

Tra il 2019 e il 2020 sono stati persi 19 milioni di visitatori e 78 milioni di euro a causa del lockdown andando a minare una tendenza positiva che continuava dal 2010 che vedeva aumentare in media di 1,7 milioni i frequentatori dei musei.

L'importanza del settore per il nostro paese viene riconosciuta anche dal PNRR (piano nazionale di ripresa e resilienza) prevedendo sussidi per la cultura e nello specifico anche i musei[10].

---

[8] ACTORS Italia, *Un modello di simulazione dell'impatto economico dei musei/istituti dell'area POIN: metodo, fonte dati, primi risultati*, 2015

[9] A. Gagliardi, *Uffizi al top per visitatori in Italia, superato il Colosseo: pesa anche l'effetto Ferragni*, Il Sole 24 Ore, 5 aprile 2022

[10] Piano Nazionale di Ripresa e Resilienza, 2021

Il suddetto documento contiene tre provvedimenti specifici:

- Il primo riguarda la digitalizzazione del patrimonio culturale italiano, stanziando circa 500 milioni e il processo si realizzerà tra il 2021 e 2026.
- Il secondo consiste nel rimuovere le barriere architettoniche presenti concedendo e facilitando l'accesso a persone con disabilità, per questo intervento è prevista una spesa di 300 milioni.
- Il terzo è destinato a migliorare l'efficienza energetica delle strutture ed un terzo del budget, 100 milioni, è riservato proprio ai musei.

1.3 Musei e covid

Ad oggi i musei si ritrovano ad affrontare diverse sfide, ma una in particolare ha causato danni notevoli all'intero settore: la pandemia causata dal Covid19.

L'obbligo di chiusura e, in seguito, quello di limitare gli accessi ha provocato enormi problemi ai musei e come citato in precedenza ha visto crollare il numero di visitatori annuali.

Secondo il primo sondaggio condotto dall'ICOM a livello mondiale nell'aprile-maggio del 2020 il 94,8% dei musei era rimasto chiuso, mentre ad un anno la situazione era migliorata passando ad un 27% di musei chiusi e 29,4% chiuso a seconda delle norme locali. Sebbene la situazione pare sia migliorata, risulta comunque che il 9,6% dello staff impiegato è stato licenziato. Inoltre, il 44% dei gestori museali intervistati dichiara di aspettarsi perdite maggiori del 50% e il 70% afferma di aver visto un calo delle visite maggiore del 50%.

Per far fronte a questa situazione i governi hanno cercato di fornire sussidi per limitare i danni anche se il 36,06% ammette di non aver ricevuto nessun supporto. Sebbene un metodo per riuscire a creare nuovi introiti fosse quello di investire in attività legate al digitale, come tour virtuali, ampliare lo store online e offrire corsi, circa il 60% non ha dato il via a nessuna alternativa. Nonostante ciò, il movimento verso il digitale da parte dei musei è in crescita.

La transizione è giustificata per far fronte alle principali paure per il futuro: ovvero riuscire ad attrarre nuovamente i visitatori persi e di perdere il contatto con la comunità locale. Per ovviare a ciò risulta necessario stabilire delle strategie in grado di comprendere le richieste e gli interessi dei potenziali clienti per questo social media e strumenti per la raccolta di dati risulta fondamentale[11].

1.4 Musei e social media

Il mantenere contatti con la comunità e attrarre nuovi visitatori è fondamentale per i musei e ora più che mai a causa del distanziamento sociale degli ultimi anni.

Gli strumenti di marketing diretto per eccellenza adatto allo scopo sono i social media, specialmente se si vuole raggiungere la fetta del mercato più giovane. L'elevato numero di utenti e la semplicità di condivisione offerta dalle piattaforme permette anche ai musei più piccoli di creare interesse: è il caso del Museum of Rural Life di Reading[12], Regno Unito, che ha postato il 9 aprile 2018 la foto di una grossa pecora su Twitter diventata poi virale con oltre 100 mila likes e 25 mila condivisioni (Figura 1.2).

[11] ICOM, *Museums, museum professionals and Covid-19: third survey*, 2021
[12] https://merl.reading.ac.uk/news-and-views/2018/04/history-behind-absolute-unit/

Figura 1.2 Tweet postato da The MERL diventato virale

Il metodo scelto per comunicare dei musei più seguiti è quello di informare e incuriosire i follower senza sembrare una pubblicità diretta a vendere biglietti, ne è un esempio il Tate[13] che introduce la sua collezione in modo leggero e informativo (Figura 1.3).

[13] https://www.tate.org.uk/visit/tate-modern

Figura 1.3 Esempio Tweet postato da Tate Modern

Questi media permettono di interagire in tempo reale con il pubblico, rispondendo alle loro domande, ai commenti e raccogliendo feedback. Inoltre, possono indirizzare gli interessati verso *long-form* media, come *blog* e *wiki*, per avvicinarli più dettagliatamente ai progetti e percorsi educazionali in modo da creare un flusso di visite al sito.

Ovviamente ciascun museo deve perfezionare la strategia più adatta a sé, poiché ciò che funziona per un'istituzione può non dare gli stessi risultati, è un procedimento dispendioso sia in termini di denaro che di tempo, ma necessario se si vuole espandere il proprio bacino di utenza.

Le piattaforme social permettono di raggiungere un vasto numero di utenti, ma l'obiettivo principale è quello di riuscire a creare una parte di utenza che interagisca attivamente e che sia incuriosita a visitare il museo, lo scopo, quindi, è quello di creare una community[14].

Il metodo più efficace per fare ciò è attuare una strategia di comunicazione ponendo il focus sulla narrazione ovvero fare storytelling. Ne è un buon esempio il Tate Modern citato in precedenza.

I consigli per sfruttare al meglio le possibilità offerte dai social sono i seguenti[15]:

- Distinguere la comunicazione tra sito web e canali social
- Rispondere tempestivamente a domande e a critiche
- Coniugare semplicità di linguaggio e l'autorevolezza dell'istituzione
- Evitare l'eccesso di comunicazione
- Raccontare cose interessanti e costituire un percorso digitale
- Definire la propria identità digitale

1.5 Musei e analisi dati

La necessità di conoscere meglio i visitatori è sempre stata una delle priorità dei musei, ciò trova conferma nel frequente utilizzo di sondaggi e richiesta di feedback periodico o al termine di una mostra o esposizione.

Sebbene il fine ultimo non sia generare profitto, i musei sono comunque istituzioni complesse che hanno degli obiettivi da raggiungere. L'espansione del digitale e i nuovi strumenti a disposizione permetto di ottenere sempre più informazioni utili per prepararsi ai nuovi trend e richieste del mercato.

Fornire la possibilità di scaricare un'applicazione dedicata alla visita, può non solo migliorare l'esperienza dei visitatori, ma fornire importanti informazioni su ciò che ha catturato maggiormente l'attenzione, come ad esempio conteggiare quali sono le opere

[14] https://www.art-museum.it/social-media-marketing-per-museo.html
[15] https://www.art-museum.it/social-media-marketing-per-museo.html

che hanno incentivato la richiesta di ulteriori informazioni, simile procedura può essere assolta anche da touch screen apposti a fianco delle opere.

È possibile anche sfruttare soluzioni meno gravose sul budget utilizzando gli strumenti offerti dalle piattaforme digitali: sondaggi, polls, Google Analytics e le statistiche consultabili dai propri account business sui social media (Figura 1.4). Per esempio, mediante le statistiche rese disponibili da Instagram si possono osservare il numero di account raggiunti in un lasso di tempo, nella sezione *interazioni* è possibile visualizzare i dati relativi alle azioni eseguite dagli utenti quando interagiscono con l'account (visite ed e-mail). Selezionando la sezione *pubblico* vengono rese disponibili informazioni riguardanti i propri follower: zona di provenienza, fascia d'età, sesso, orari e giorni nei quali sono più attivi.

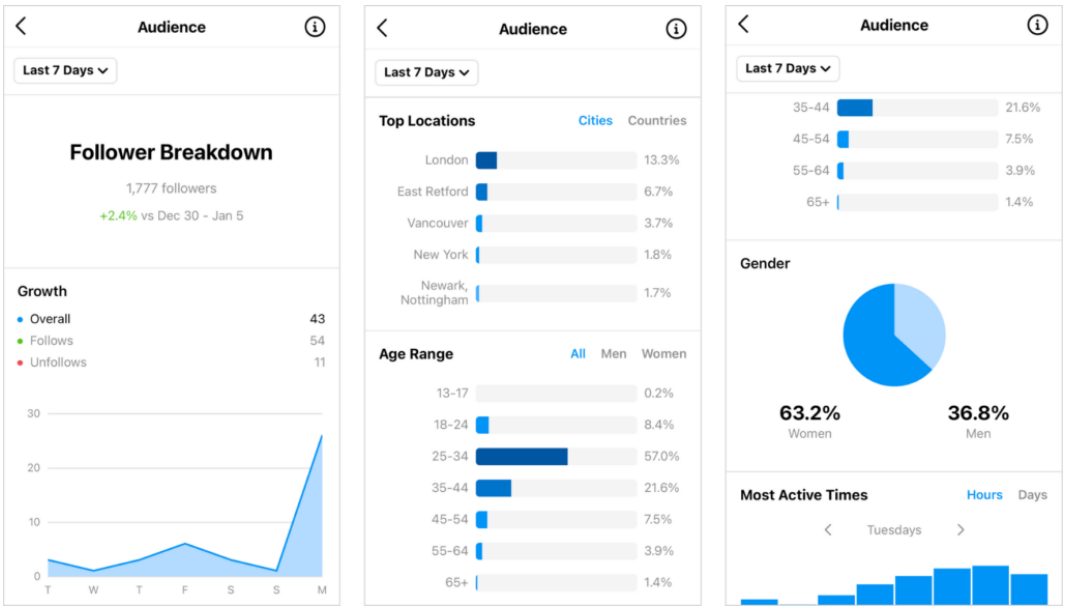


Figura 1.4 Esempio dei dati disponibili tramite Instagram

La quantità di dati facilmente reperibili derivanti da passate esibizioni, dai social media e dalla concorrenza è enorme, così ogni museo ha l'opportunità di cimentarsi nell'analisi di dati e di conseguenza attrezzarsi al meglio per affrontare il futuro.

1.5.1 Analisi spostamenti visitatori nel Louvre

Oltre ai dati demografici ottenibili tramite sondaggi è possibile implementare sistemi di tracciamento elettronici per monitorare gli spostamenti dei visitatori e le aree di interesse. Per esempio, uno studio condotto presso il Louvre e riprodotto similmente in altri musei, si è posto l'obiettivo di studiare i movimenti all'interno dell'area ed evitare il congestionamento. Per questo obiettivo sono stati utilizzati sensori bluetooth posti nel percorso principale del Louvre, che va dalla piramide alla Venere di Milo. Mediante questi strumenti è stato possibile captare informazioni sul passaggio dei visitatori che mantenevano i bluetooth acceso sul loro dispositivo ed è risultato che le due categorie di frequentatori (quelli con una lunga permanenza e quelli a breve) si comportano allo stesso modo: ovvero durante la visita tendono a soffermarsi su alcuni luoghi chiave e che l'unica differenza è il tempo trascorso in quella sezione[16].

1.5.2 Analisi reazioni emotive Bologna Musei

Un altro metodo che si sta sperimentando è quello di analizzare la reazione emotiva delle persone di fronte ad un'opera mediante un sensore in grado di riconoscere le espressioni facciali. Ne è un esempio ShareArt, sviluppato da ENEA (L'agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico) che dal 15 giugno 2021 ha dato il via, assieme all'Istituzione Bologna Musei, ad un progetto in grado di osservare le reazioni emotive dei visitatori. ShareArt è composto da diversi dispositivi aventi una telecamera che raccolgono le informazioni che vengono poi inviate ad un programma in grado di analizzarle[17].

---

[16] Yoshimura, Y., Sobolevsky, S., Ratti, C., Girardin, F., Carrascal, J.P., Blat, J. and Sinatra, R., 2014. An analysis of visitors' behavior in the Louvre Museum: A study using Bluetooth data. Environment and Planning B: Planning and Design, 41(6), pp.1113-1131.

[17] https://www.enea.it/it/Stampa/news/beni-culturali-al-via-il-progetto-shareart-per-misurare-il-gradimento-delle-opere-d2019arte-con-l2019intelligenza-artificiale

2 Indagine sul comportamento dei visitatori analizzando i post di Instagram

2.1 Introduzione

Come accennato nel capitolo precedente, oggi i social media sono una delle forme di comunicazione più utilizzate al mondo e un potente strumento di marketing diretto. Ad ottobre 2021 si contano circa 4,55 miliardi (datareportal, 2021) di utenti attivi. Tra le piattaforme più popolari vi sono Facebook, Twitter, Instagram, WhatsApp e TikTok (similarweb, 2022).

La diffusione capillare dei social media porta le persone a scegliere mete e destinazioni turistiche apposite con l'obiettivo di poter condividere sul web le proprie esperienze culturali (Dudareva, 2014; Phan and Macias, 2018).

Prendendo in considerazione i musei ci si può chiedere cosa possa portare un possibile interessato a visitarlo e cosa poi, di conseguenza, posterà per raccontare la sua esperienza.

L'impatto visivo è la caratteristica distintiva delle esposizioni museali, dunque, il mezzo principale con il quale raccontare la propria visita sono le foto. Così Instagram, il social media di image sharing più utilizzato (similarweb, 2022), può rivelarsi estremamente utile per poter comprendere cosa i visitatori fotografano più frequentemente e ciò che sono più interessati ad immortalare.

In questo capitolo verranno prese in considerazione le informazioni e dati contenuti nel paper di Rhee, B. A., Pianzola, F., and Choi, G. T. 2020 Analyzing the museum experience through the lens of Instagram[18] e si illustreranno i procedimenti e i risultati ottenuti. La ricerca si articola su due punti:

- Analisi degli hashtag utilizzati nei post di Instagram relativi ai musei presi in considerazione per osservare le informazioni ricavabili da essi.
- Analisi delle foto postate su Instagram relative ai musei presi in considerazione per poter comprendere ciò che i visitatori fotografano e

[18] Rhee, B. A., Pianzola, F., and Choi, G. T. 2020 Analyzing the museum experience through the lens of Instagram posts

condividono online e se ci sono differenze a seconda della tipologia di museo.

2.2 Ricerche precedenti

L'analisi delle foto di Instagram da parte dei musei è un campo ancora parzialmente inesplorato, probabilmente per via della mancanza di strumenti idonei allo scopo. Sebbene la ricerca sia ancora nelle fasi iniziali, dagli studi condotti finora possono essere tratte alcune conclusioni. I musei utilizzano i social media come strumento per creare dialogo con la comunità (Capriotti and Kuklinski, 2012). I visitatori usano Instagram per combinare l'esperienza museale con la loro narrativa personale in un processo di creazione della propria immagine, ricordo, ispirazione estetica e gioco (Weilenmann et al., 2013; Budge and Burness, 2018; Budge, 2018; Suess, 2018; Chlebus-Grudzień, 2018; Villaespesa and Wowkowych, 2020).

Una ricerca condotta presso una galleria australiana ha riportato che i post di Instagram facilitano la promozione delle esibizioni (Suess, 2014). Un ulteriore studio sui post riguardanti musei di scienze ha mostrato come spesso manchino hashtag o immagini poco legate alla scienza, mentre un terzo di essi non aveva alcun legame con contenuti scientifici (Jarreau et al., 2019), rendendo così difficile capire come effettivamente i visitatori rielaborino la propria visita al museo.

Una ricerca simile a quella che viene presa in esame in questa tesi è quella condotta dall'Australian Museum of Contemporary Art (Budge and Burness 2018). Per comprendere il comportamento delle persone sono state analizzate 390 immagini e suddivise in categorie: solo oggetti, persone ed oggetti, solo persone e altri. Ne è risultato che i frequentatori postano più foto relative agli oggetti piuttosto che selfie (Budge and Burness, 2018; Budge, 2017; Villaespesa and Wowkowych, 2020).

La ricerca presa in esame per questa tesi si è posta l'obiettivo di comprendere cosa i visitatori decidono di condividere riguardo alla loro esperienza nei musei, analizzando immagini e hashtag e di fornire uno strumento ed una metodologia per poter sfruttare le foto dei musei per trarre in formazioni da utilizzare per pianificare strategie.

2.3 Instagram

Il social media preso in considerazione per la raccolta di immagini è Instagram, in quanto principale piattaforma di *image sharing*. È stato creato nel 2010 da Kevin Systrom e Mike Krieger e in seguito acquisito, nel 2012, da Facebook Inc., ora Meta Platforms. Il social network permette agli utenti di postare contenuti multimediali organizzandoli con *hashtag*[19] e *geotagging*[20]. Sebbene i dati riguardanti al numero di utenti non siano più stati dichiarati dal 2018, stando a CNBC, nel 2021, Instagram ha raggiunto i 2 miliardi di utenti attivi mensili, inoltre nel 2021 è stata la quarta app più scaricata[21].

2.4 Dati considerati

La ricerca di riferimento ha preso in considerazione otto musei, due per ogni nazione considerata (Francia, Regno Unito, Stati Uniti, Corea del Sud), uno di arte generale e l'altro di arte moderna per comprendere se ci sono differenze a seconda della tipologia. I musei considerati, esclusi quelli coreani, sono tra quelli con maggiore presenza su Instagram (Tabella 2.1):

- British Museum (nominato anche con l'abbreviazione BM), UK
- Tate Modern (nominato anche con l'abbreviazione TM), UK
- Metropolitan Museum of Art (nominato anche con l'abbreviazione Met), USA
- Museum of Modern Art (nominato anche con l'abbreviazione MoMA), USA
- Louvre Museum (nominato anche con l'abbreviazione LM), FR
- Pompidou Centre (nominato anche con l'abbreviazione PC), FR
- National Museum of Korea (nominato anche con l'abbreviazione NMK), KR
- National Museum of Modern and Contemporary Art Seoul (nominato anche con l'abbreviazione MMCA), KR

Nella riproposizione i musei coreani non verranno presi in considerazione per motivi linguistici. Inoltre, per evitare *bias* è stato consultato il regolamento riguardo alla

[19] Tag metadato, utilizzato su alcuni social media e servizi web, come aggregatore tematico, il quale rende più semplice per gli utenti trovare contenuti su un argomento specifico.
[20] Processo che permette di inserire coordinate geografiche ai media basandosi sulla posizione del dispositivo.
[21] https://www.cnbc.com/2021/12/14/instagram-surpasses-2-billion-monthly-users.html

possibilità di effettuare fotografie per ogni museo e si può concludere che hanno tutti un regolamento molto simile, ovvero le foto sono ammesse senza flash per uso personale, salvo quando espressamente vietato.

Il primo passo è raccogliere informazioni dai post di Instagram e lo strumento utilizzato per la scopo è 4kStogram, uno strumento online, usato per effettuare backup del proprio account ma che permette anche di cercare e scaricare sul pc foto, reels, storie e tag di altri account ricercandole per luogo o hashtag[22].

Per ogni museo sono state raccolte 10,000 immagini, postate tra il 22 settembre 2019 e il 2 novembre 2019, provenienti da un numero di post differente in quanto è possibile caricare più foto in ognuno, con l'hashtag contenente il nome del museo (e.g. #louvre, #louvremuseum, ecc.…).

Ovviamente data la natura delle informazioni non è possibile catalogare demograficamente la popolazione creatrice dei contenuti, ma si possono solo considerare le caratteristiche generali degli utilizzatori della piattaforma, sapendo che i viaggi sono l'interesse più seguito dagli utenti (43%) (Facebook IQ, 2019) e la fascia di età più rappresentata è quella tra i 18 e i 34 anni, equamente distribuita tra uomini e donne (Kemp, 2019).

2.5 Analisi hashtag

La prima parte della ricerca presa in esame si propone di comprendere quali informazioni si possono ricavare dagli hashtag dei post. Sono stati presi in considerazione 1800 post, rappresentante il numero minimo di elementi disponibili per un solo museo e rimuovendo l'hashtag con il nome del museo, mantenendo, però, le parole derivate, dato che la ridondanza rappresenta un'informazione utile alla ricerca, per esempio, il nome ufficiale #louvremuseum non è stato considerato, ma #louvre sì. Successivamente sono state selezionate le 50 hashtag più utilizzate e suddivise in informative (relative a informazioni circa la località, la mostra, o l'arte in generale), partecipative (quando un determinato user propone ai suoi follower di postare foto idonee per poter attribuire un

[22] https://www.4kdownload.com/-ad0p9/stogram

determinato hashtag) ed exposure (utilizzate per aumentare la popolarità del proprio post) (Kim, 2016; Oh et al., 2016; Rauschnabel et al., 2019; Park et al., 2019).

Tabella 2.1 Numero visitatori e followers musei considerati

Nome del museo	Numero di visitatori 2021	Numero di followers 2022
LM	9.720.260	4.908.873
PC	3.800.000	1.194.404
MET	6.115.881	4.163.124
MoMA	2.805.659	5.621.337
BM	5.575.946	2.085.611
TM	5.304.710	4.180.186

Dalla ricerca si osserva che per i musei coreani vengono utilizzate molti meno hashtag rispetto agli altri, ma caricano più immagini per ogni post (Tabella 2.2). Considerando le suddivisioni presentate in precedenza, pare che le categorie riguardanti hashtag che descrivono l'arte in generale siano quelle più utilizzate, seguite da quelle su informazioni locali, per poi chiudere con quelle relative alla sede espositiva. Fanno eccezione il Met e il Louvre con relativamente pochi hashtag riguardanti l'arte, ma molti legati a informazioni locali (Tabella 2.3).

Si può notare, che per i musei occidentali viene utilizzato raramente il nome degli artisti, mentre il termine legato alla corrente artistica è usata con maggiore frequenza per i musei d'arte. Per finire, i nomi specifici degli artefatti sono usati molto di rado.

Tabella 2.2 Analisi quantitativa hashtag per i post

Nome museo	Numero di post	Media immagini per post	Numero totale di hashtag (n=1800)	Media hashtag per post (n=1800)
LM	4,106	2.4	14,358	8.0
PC	4,615	2.2	11,158	6.2
MET	3,366	3	10,554	5.9
MoMA	4,694	2.1	16,959	9.4
BM	3,231	3.1	11,019	6.1
TM	3,693	2.7	11,035	6.1

NMK	2,010	5	6,672	3.7
MMCA	1,848	5.4	7,599	4.2

Tabella 2.3 Riassunto categorizzazione hashtag

Risultati categorizzazione degli hashtag

Nome museo	Tipo informativo			Exposure (%)	Totali di hashtag usati per 50+ volte	Hashtag più utilizzata (n)
	Spazio espositivo (%)	Informazioni locali (%)	Arte in generale (%)			
LM	25.6	40.2	18.0	16.2	7,722	#paris (1,296)
PC	12.1	33.7	51.0	3.20	4,724	#paris (857)
MET	29.1	38.7	28.1	4.10	4,982	#nyc (581)
MoMA	8.70	13.9	71.4	6.00	8,642	#art (611)
BM	10.9	38.6	39.6	10.9	4,363	#london (90)
TM	16.9	29.3	51.0	2.80	4,507	#london (751)
NMK	16.8	25.1	37.7	20.4	1,997	#sejong (1,446)
MMCA	29.1	21.7	37.0	12.2	3,267	#mmca (382)

2.6 Analisi immagini

Una volta completata la raccolta è necessario descrivere e catalogare le foto acquisite, per questo scopo è stato utilizzato Google Vision API. Si tratta di un programma istruito tramite il machine learning in grado di riconoscere gli elementi presenti in un'immagine mediante l'assegnazione di tag con precisione decrescente. L'algoritmo per ogni soggetto riconosciuto assegna un punteggio indice dell'affidabilità del riconoscimento. Il procedimento è piuttosto veloce e semplice, come è possibile vedere nella Figura 2.1, caricando una foto della parte frontale del Louvre il software analizza l'immagine assegnando tag, visibili accanto allo scatto, basandosi su ciò che riconosce e con che precisione l'elemento è effettivamente raffigurato (e.g. il software è "sicuro" al 97% che ci sia dell'acqua).

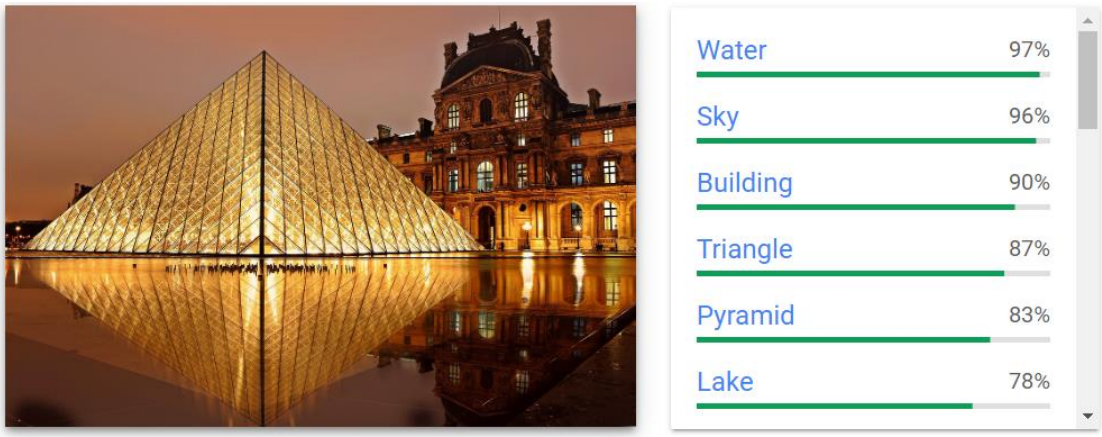


Figura 2.1 Attribuzione tag di Google Vision

Per questa ricerca non sono stati imposti filtri sul punteggio minimo, ma, sono state considerate le prime 10 etichette.

Tabella 2.4 Riassunto statistiche di attribuzione etichette di Google Vision

Nome museo	Media etichette assegnate	Mediana etichette assegnate
MoMA	8.776	10
Tate	8.409	10
Met	8.821	10
Louvre	9.367	10
CentrePom	8.324	10
BM	8.668	10

Si può osservare che per ogni museo la mediana delle etichette assegnate è stata il massimo stabilito (Tabella 2.4).

In seguito, è stato utilizzato Word2vec[23], un insieme di modelli utilizzati nel *word embedding* in grado di partire da un *corpus*, un insieme di testi organizzati e selezionati, e generare un vettore per ciascuna parola rappresentandola così come un punto nello spazio e trovandosi vicina a parole con una certa similarità semantica. Nello specifico

[23] Mikolov, T., Chen, K., Corrado, G. and Dean, J., 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Word2vec rappresenta le parole in vettori con i quali utilizzando la similarità del coseno è possibile indicare il livello di similarità semantica tra le parole rappresentate dai vettori.

L'output ottenuto è stato ridotto a due dimensioni per semplificarne l'osservazione. Partendo da questa nuova disposizione sono stati individuati dei raggruppamenti suddivisibili in *body*, *food*, *clothes*, *music*, *nature*, *interior*, *architecture*, *museum*, *animals* e *sport*. Da questi sono stati considerati solo quelli rilevanti per un'analisi da parte dei musei creando così le categorie definitive (Figura 2.2):

- Art exhibition (es. eventi, performance)
- Artifact (es. sculture, vasi, quadri)
- Architecture (es. edifici, parte di essi, interni)
- Selfie (tratti del volto)
- Food (cibo)
- Human body (es. parti del corpo escluso il volto, persone)
- Landscape (es. elementi naturali, spazi all'aperto)

Figura 2.2 Esempi di suddivisione categorie Art exhibition (a), Artifact (b), Architecture (c), Food (d), Human body (e), Landscape (f), manca l'esempio di un selfie per motivi di privacy.

Fatto ciò, è stato utilizzato nuovamente Word2vec per trovare le 50 etichette più simili per ciascuna categoria e manualmente sono state eliminate quelle ambigue e

mantenute solamente quelle rilevanti. In seguito, sono state sostituite nel dataset le etichette rielaborate ed eliminate quelle che non avevano alcuna corrispondenza. Per finire è stata eseguita un'analisi di frequenza sull'etichette di ogni museo per osservare le categorie più fotografate dai visitatori.

I risultati ottenuti di seguito sono frutto della rielaborazione eseguita partendo dai dati ottenuti dalle etichette prodotte dal software Google Vision API e facendo una *cross reference* con le categorie illustrate precedentemente. Per esempio, le tag "*coffin", "chariot" e "astronomical object"* assegnate da Google Vision e poi raggruppate per aree semantiche da Word2vec sono state poi riunite in un'unica categoria: "*Artifact"*, in quanto tutti sono degli artefatti.

Iniziando dal dataset contenente le tag generate da Google Vision si procede mediante R a considerare le tag di ciascuna immagine come frasi per poi creare dei token delle parole per poter poi effettuare un'analisi quantitativa e per poter poi produrre grafici per illustrare i risultati.

Per eseguire questa ricerca è stata usata la versione 4.2.1 di R[24] e principalmente il pacchetto *tidytext* versione 0.3.4[25]

Per ciascun museo è stato generato un ranking delle etichette più utilizzate, considerando solo le parole con almeno 1000 occorrenze, e in seguito viene proposto un *wordcloud*[26] utilizzando il pacchetto *wordcloud2* per rappresentare i risultati (Figure dalla 2.3 alla 2.14).

[24] https://cran.r-project.org/
[25] https://cran.r-project.org/web/packages/tidytext/index.html
[26] Tipo di grafico utilizzato per visualizzare la frequenza con la quale una parola appare nel testo, maggiore è la dimensione, maggiore è la frequenza.

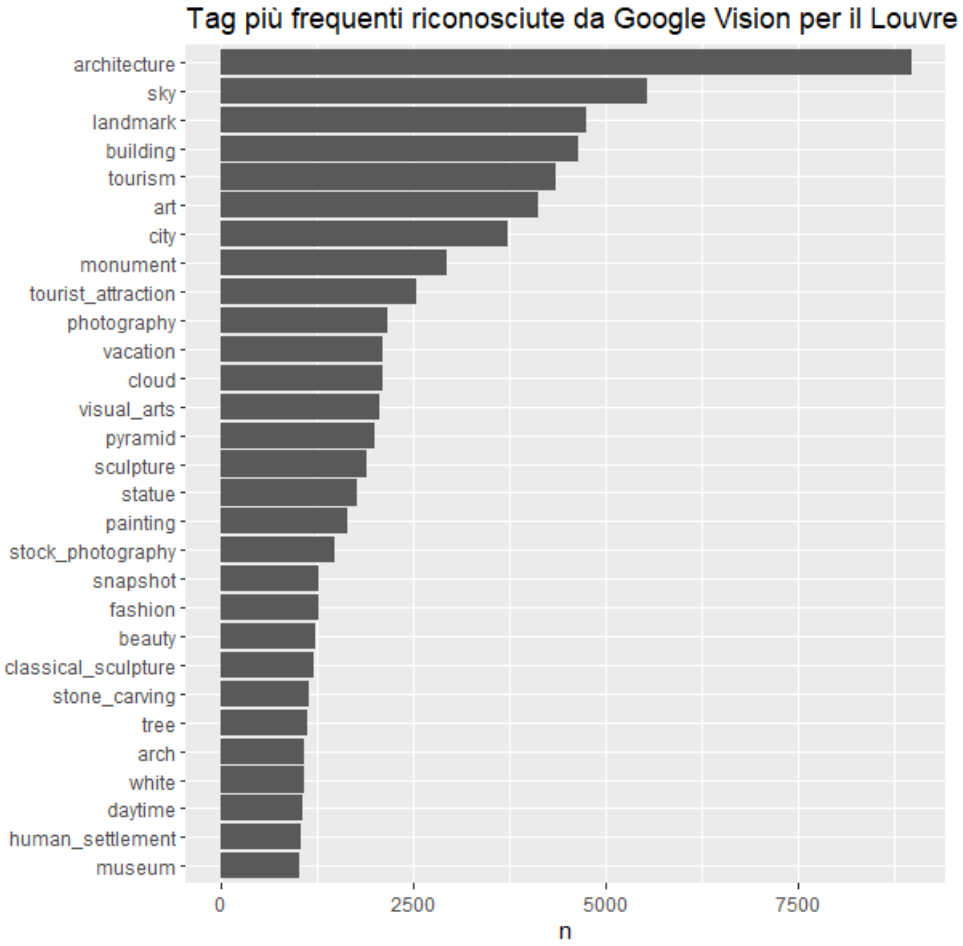


Figura 2.3 Grafico rappresentante le etichette più frequenti per i post legati al Louvre

Figura 2.4 Wordcloud rappresentante le etichette più frequenti per i post legati al Louvre

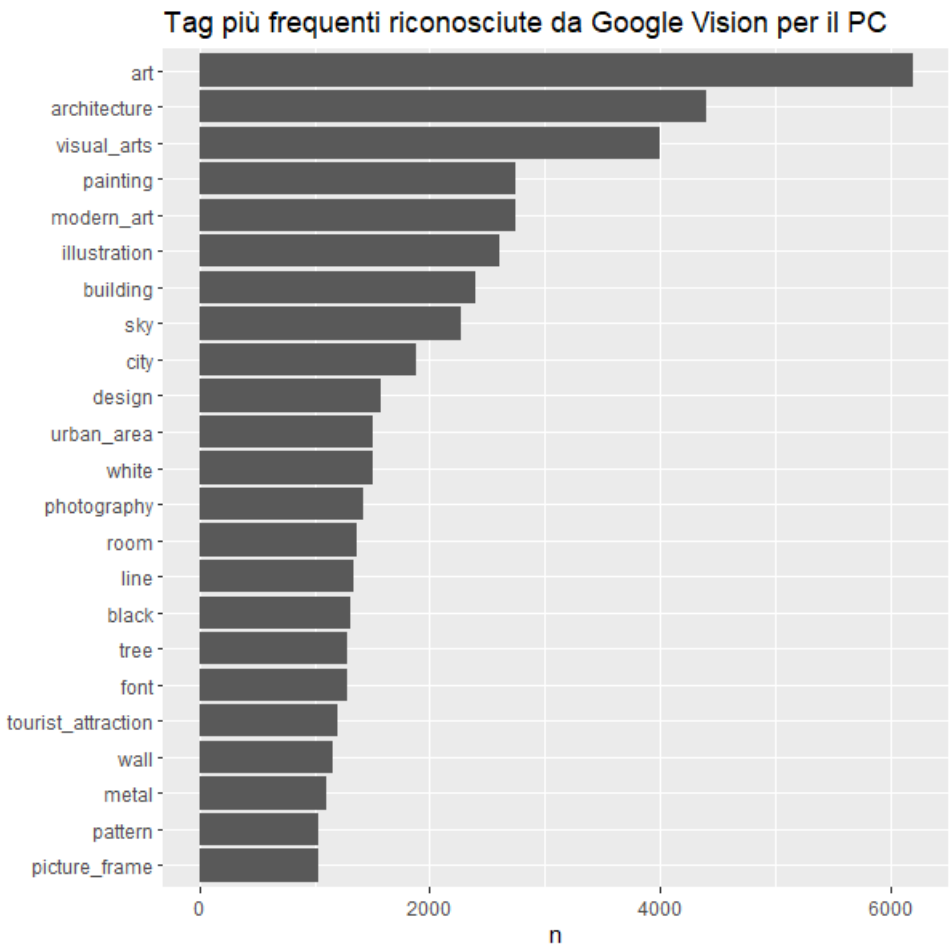


Figura 2.5 Grafico rappresentante le etichette più frequenti per i post legati al Pompidou

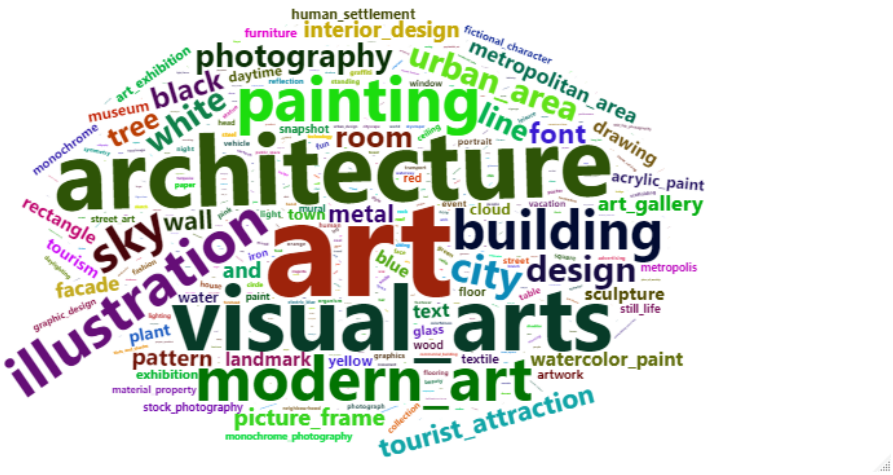


Figura 2.6 Wordcloud rappresentante le etichette più frequenti per i post legati al Pompidou

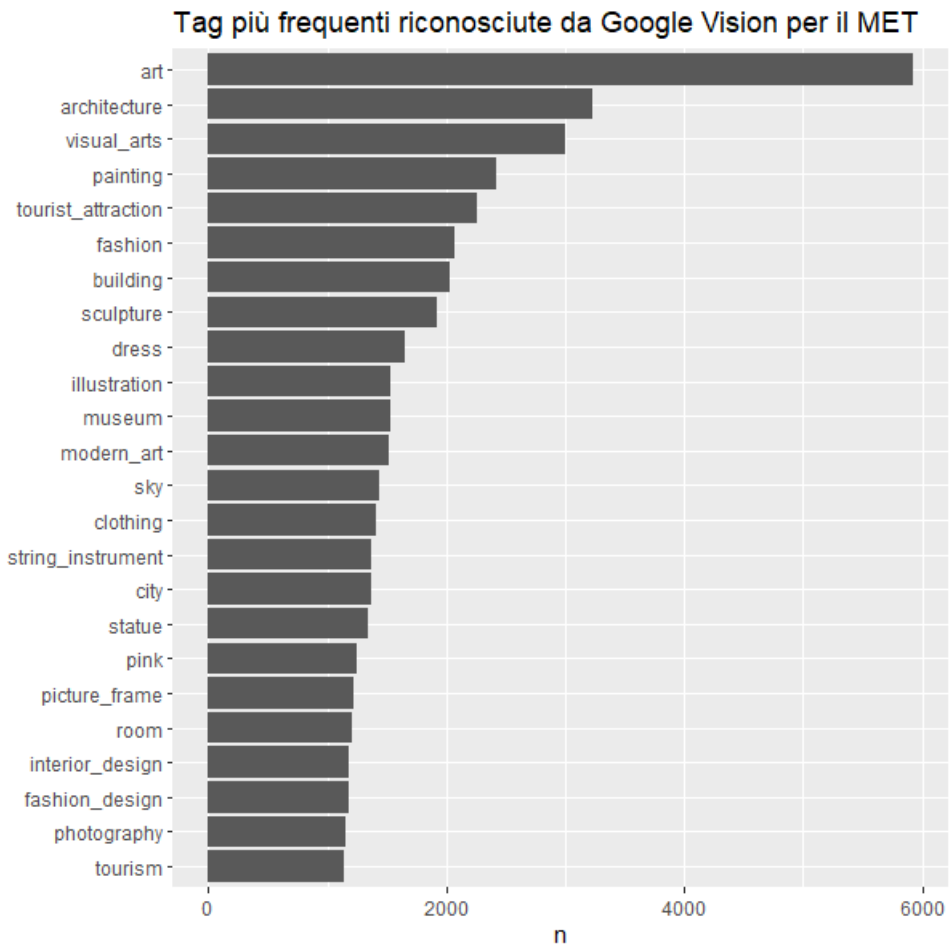


Figura 2.7 Grafico rappresentante le etichette più frequenti per i post legati al Met

Figura 2.8 Wordcloud rappresentante le etichette più frequenti per i post legati al Met

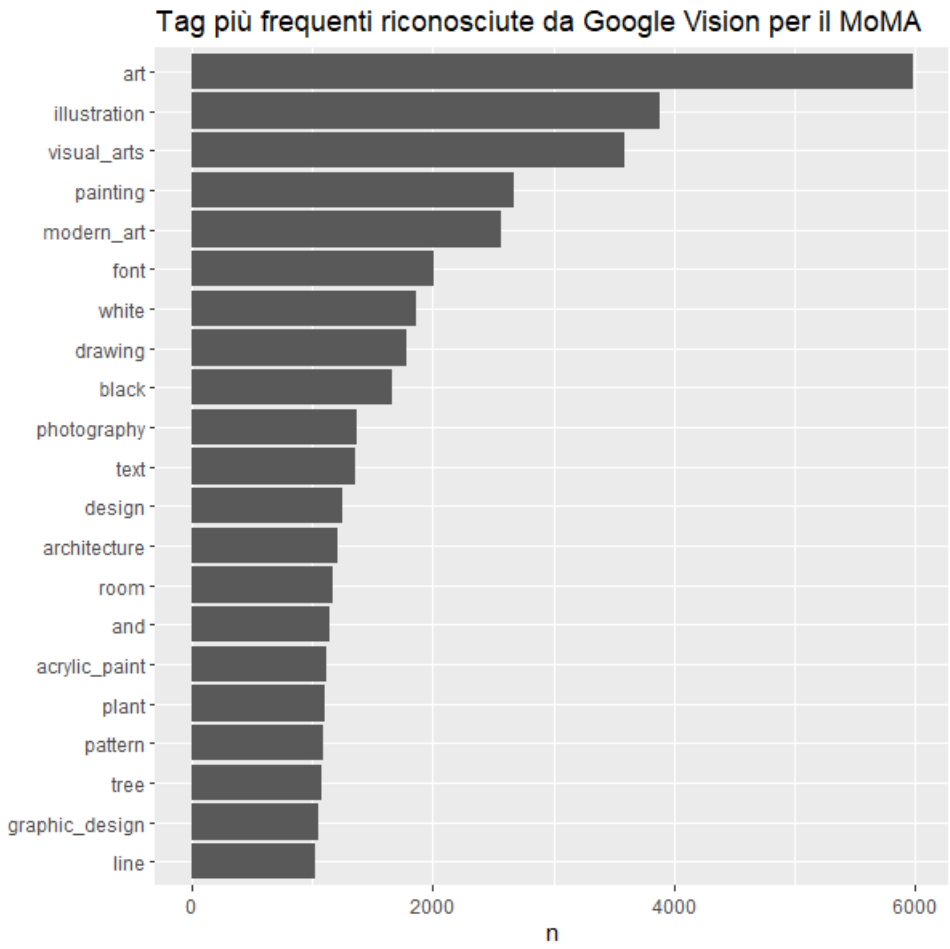


Figura 2.9 Grafico rappresentante le etichette più frequenti per i post legati al MoMA

Figura 2.10 Wordcloud rappresentante le etichette più frequenti per i post legati al MoMA

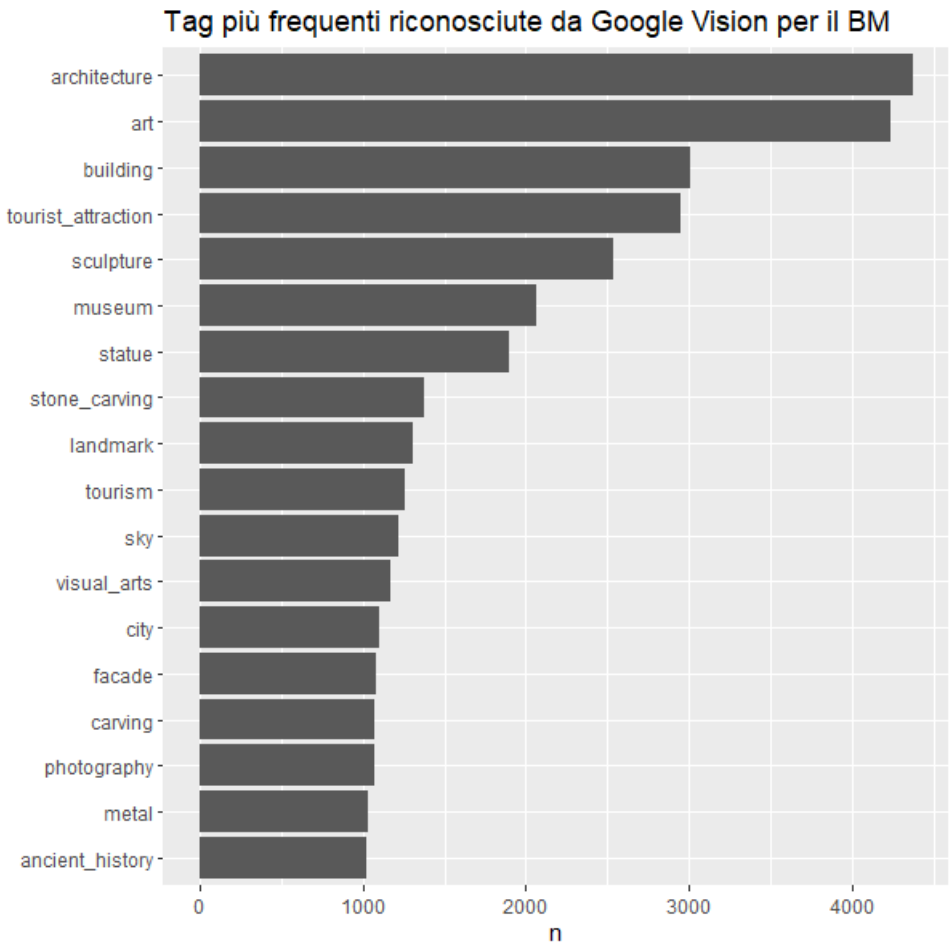


Figura 2.11 Grafico rappresentante le etichette più frequenti per i post legati al BM

Figura 2.12 Wordcloud rappresentante le etichette più frequenti per i post legati al BM

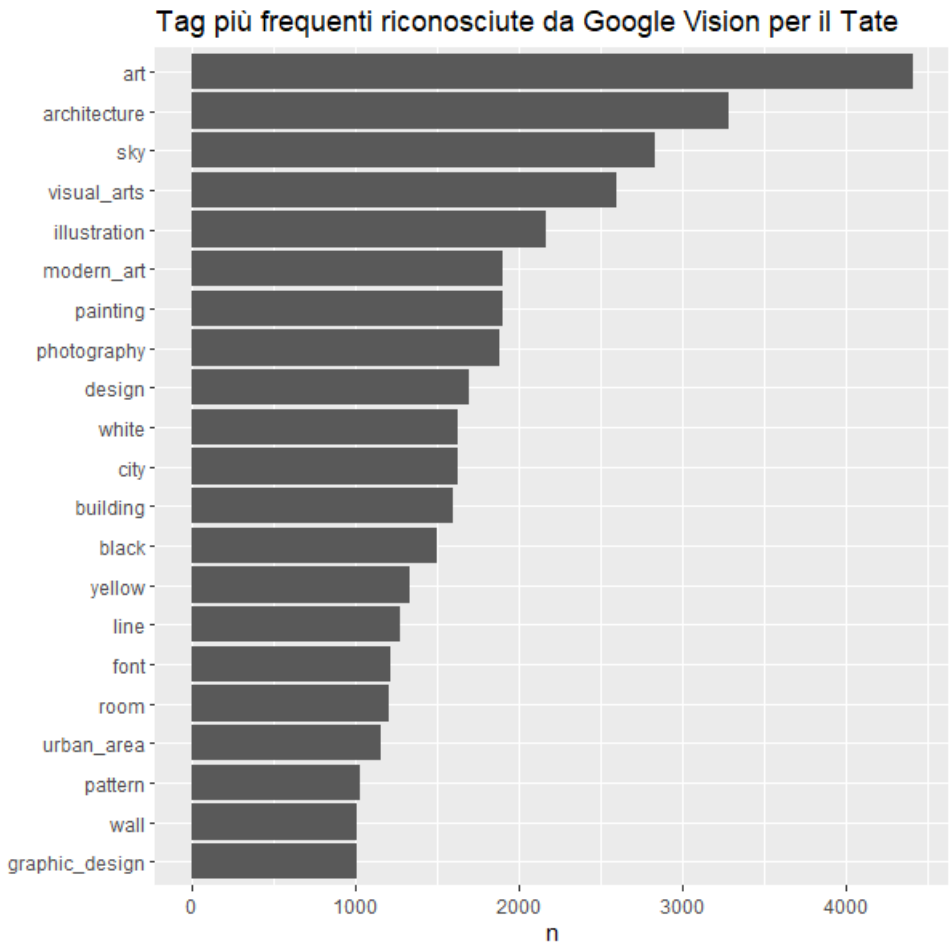


Figura 2.13 Grafico rappresentante le etichette più frequenti per i post legati al Tate

Figura 2.14 Wordcloud rappresentante le etichette più frequenti per i post legati al Louvre

2.7 Sintesi analisi delle immagini

Complessivamente, a seconda della tipologia di museo, si possono osservare differenze nella frequenza di determinate categorie piuttosto che altre (Figura 2.15).

Per i musei di arte moderna si può notare come la categoria *art_exhibition* è molto più frequente (tra il 25,64% del Tate Modern e il 31,48% del MoMA) la differenza possibilmente è dovuta al fatto che i musei generalisti presentano più quadri o opere che rientrano nella categoria *Artifact*. Per i musei di arte in generale, invece, sono l'architettura, gli spazi interni e gli artefatti ad essere più fotografati. Fa eccezione il Louvre dove le immagini relative all'architettura sono preponderanti (il 42,5%), probabilmente dovute alla popolarità del palazzo e della Piramide del Louvre. Lo stesso vale per il British Museum con un elevato numero di foto che mostrano la struttura (il 36,43%) dovute alla popolarità della facciata. L'opposto vale per il MoMA la cui struttura è la meno immortalata (12%) magari causata dalla mancanza di impatto visivo dell'edificio.

In ogni caso la frequenza di selfie è piuttosto bassa e simile ovunque (tra il 4,9% del Pompidou e il 7,67% del Louvre) indicando un maggior interesse per ciò che esposto piuttosto che per sé stessi. Fa eccezione il MoMA che presenta un buon numero di selfie (10,38%) e assieme alla frequenza di immagine con l'intera persona (9,48%) potrebbe indicare che i visitatori del MoMA abbiano un interesse a fotografare sé stessi o altre persone all'interno del museo.

In generale dai dati e dalla relativa analisi si potrebbe sostenere che i visitatori tendono a fotografare la struttura o come l'oggetto è predisposto nello spazio museale, senza focalizzarsi sul singolo artefatto. Il fatto che la categoria *Landscape* è piuttosto frequente indica come vi è un vero e proprio interesse ad immortalare il museo, forse per la riconoscibilità e per generare interesse sui social.

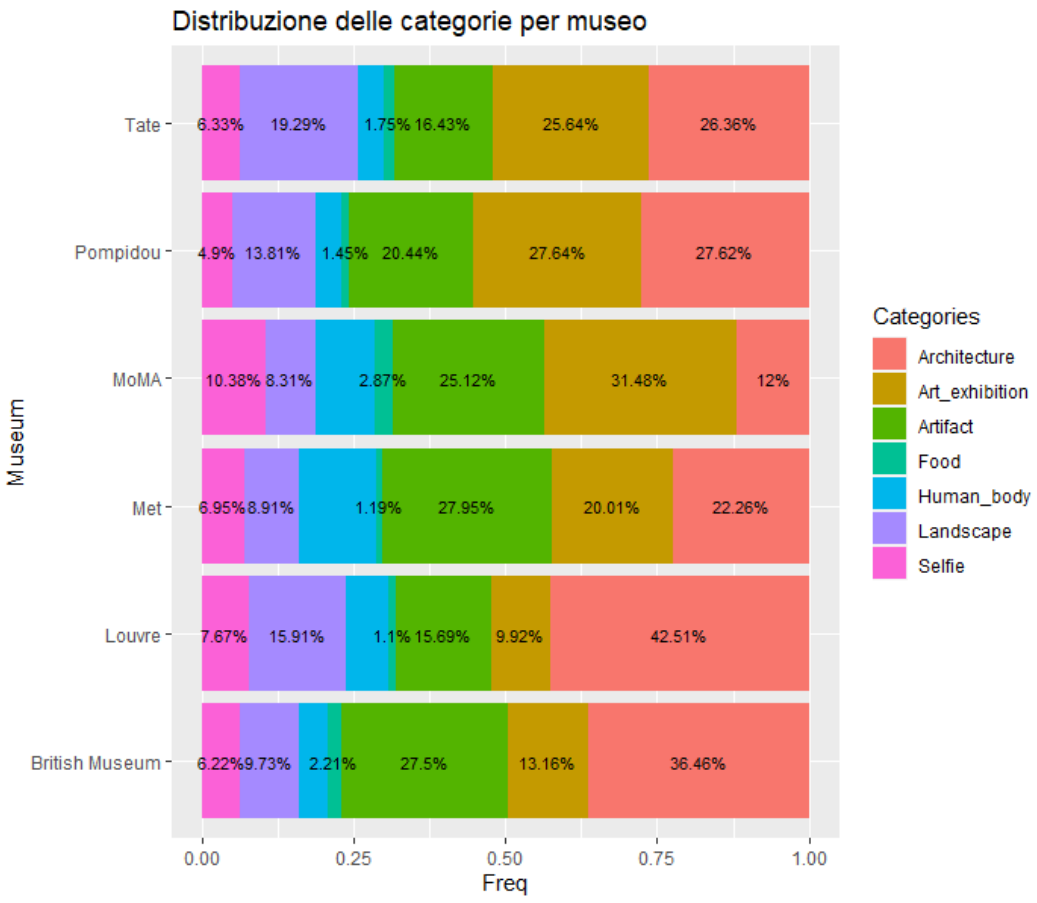


Figura 2.15 Grafico rappresentante la distribuzione delle categorie per ciascun museo

3 Analisi recensioni TripAdvisor

3.1 Perché analizzare le recensioni

Per rafforzare la ricerca condotta nel capitolo precedente si ritiene opportuno effettuare un'analisi sulle opinioni dei visitatori, le quali, al giorno d'oggi sono facilmente reperibili su Internet.

Per lo scopo si è scelto di consultare le recensioni postate su TripAdvisor, in quanto più specifico per l'attività turistica e prendendo in considerazione il MoMA di New York.

Riuscire ad avere una visione generale delle opinioni dei clienti non è semplice, in quanto oltre il punteggio espresso da 1 a 5, che l'utente aggiunge alla sua recensione, dando un voto rappresentato dai pallini, con il quale 1 rappresenta il voto peggiore e 5 il migliore, non vi è un modo per conoscere effettivamente ciò che i visitatori esprimono, se non leggendo ogni singola recensione, ma ciò sarebbe troppo dispendioso in termini di tempo. Per questo la ricerca propone una *sentiment analysis, un'analisi sui bigrammi* e un'analisi utilizzando la statistica *tf-idf* eseguite in R, utilizzando sempre il pacchetto *tidytext*, sulle recensioni ottenute tramite *web scraping*.

Gli obiettivi di quest'analisi sono:

- Osservare e quantificare con la *sentiment analysis* la risposta emotiva dei visitatori.
- Comprendere e visualizzare le relazioni tra le parole utilizzate nelle recensioni.
- Per ciascun anno comprendere quali sono state le parole più significative mediante la statistica tf-idf.

3.2 TripAdvisor

tripadvisor.com è il più grande sito dedicato ai viaggi nel mondo (similarweb, 2022) e ad oggi presenta oltre un miliardo di recensioni per quasi 8 milioni di attività. Il sito è disponibile in 44 mercati e tradotto in 22 lingue.

Il sito è il prodotto di punta di TripAdvisor Inc., un'impresa statunitense operante nel settore del turismo come agenzia di viaggio online. Opera tramite il sito web e

applicazione che permette di postare recensioni (Figura 3.1) e confrontare prezzi e offre anche la possibilità di effettuare prenotazioni.

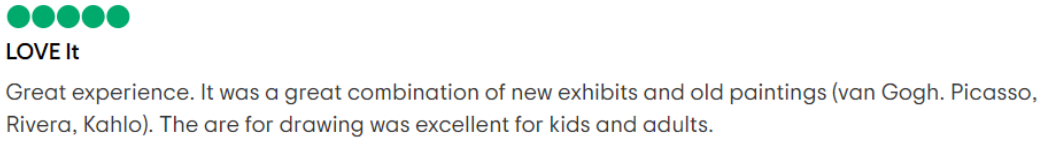


Figura 3.1 Esempio Recensione di TripAdvisor

3.3 Importanza delle recensioni

Le recensioni sono un potente strumento per i visitatori e un'essenziale fonte di informazioni per i gestori. Specialmente oggi dove è facile esprimere le proprie opinioni ed è altrettanto semplice consultarle. In ambito turistico i pareri degli utenti possono essere fondamentali nello scegliere di visitare una destinazione oppure no.

Sul web esistono numerosi modi e siti dove poter esprimere la propria opinione, dai più generici come Google Maps e Trustpilot a quelli più specializzati, ad esempio TripAdvisor per il settore del turismo.

Secondo un sondaggio svolto da Brightlocal nel 2021 su un campione di 1124 consumatori statunitensi, si riporta che il 98% degli intervistati consulta recensioni prima di effettuare acquisti o usufruire dei servizi di imprese locali. Inoltre, il 43% ha ammesso di consultare regolarmente le recensioni prima di interagire con le attività precedentemente accennate, mentre il 34% lo fa sempre[27].

Anche la mole di recensioni postate periodicamente è rilevante, nel 2015 Trustpilot sosteneva di ricevere più di mezzo milione di recensioni al mese (Forbes 2015), mentre nel 2020 su TripAdvisor ne sono state scritte oltre 26 milioni[28] di cui 4 milioni per la categoria attrazioni, esperienze e attività, tra cui sono inclusi anche i musei.

[27] https://www.brightlocal.com/research/local-consumer-review-survey/
[28] https://www.tripadvisor.com/TransparencyReport2021

Si può affermare che le recensioni abbiano un effetto sulle attività interessate, e che esse siano positive o negative, forniscono informazioni utili per analizzare il mercato, la concorrenza e per migliorare la propria offerta[29].

3.3.1 Recensioni false

La facilità con cui si può diffondere la propria opinione è un'arma a doppio taglio, in quanto è possibile abusarne e postare recensioni false o non realmente indicative dell'effettiva esperienza che si è vissuta.

In gioco ci sono molti interessi economici visto la sempre più grande importanza che si dà ad esse e stando al CMA influenzano circa 23 miliardi di sterline di spesa da parte dei consumatori ogni anno (Competition and Market authority, 2015)[30]. Non ci si ferma solo ad acquistare recensioni positive, ma anche a sopprimere quelle negative oppure a inondare di recensioni negative una determinata attività con lo scopo di danneggiarla.

Un'indagine condotta da Which? , un'organizzazione inglese con lo scopo di facilitare i consumatori ad effettuare scelte più consapevoli recensendo prodotti, servizi e informandoli sui diritti dei consumatori, nel 2019 su TripAdvisor sui principali Hotel di ogni regione considerata ha avuto come risultato che una recensione su sette risulta sospetta o molto probabilmente falsa[31] sebbene la ricerca sia stata criticata per il suo metodo semplicistico, alcuni degli hotel considerati erano stati effettivamente segnalati da TripAdvisor per la presenza di false recensioni.

Se si estende l'indagine anche sugli apparecchi elettronici più popolari su Amazon risulta che il 50,7% della top 10 delle cuffie Bluetooth e il 58,2% delle casse Bluetooth sono false o altamente sospette[32].

---

[29] R. Baggio, *Recensioni online: una rassegna della ricerca scientifica,* 2010
[30] CMA, *Online reviews and endorsement*, 2015
[31] https://www.which.co.uk/news/article/fake-tripadvisor-reviews-push-worlds-best-hotels-up-the-rankings-ayEVG8p3qiXY
[32] https://www.brightlocal.com/learn/review-management/fake-reviews/why-are-fake-reviews-a-problem/#

Per contrastare il fenomeno TripAdvisor, per esempio, opera con una squadra di esperti di frodi e lavora attivamente per rimuovere recensioni fraudolente dalla piattaforma, nel 2020 ne sono state rimosse circa 1 milione di cui il 67,1% prima di essere pubblicate[33].

3.4 Analisi delle recensioni

Per la ricerca sono state raccolte recensioni rivolte al MoMA a partire dal 20 ottobre 2014 al 2 luglio 2022 per un totale di 7,186 recensioni scritte in lingua inglese, scelta per la maggiore disponibilità di strumenti appositi per questo idioma.

Per eseguire lo *scraping* è stata utilizzata un'estensione web chiamata Web Scraper disponibile su webscraper.io che offre una versione gratuita con funzionalità di base, ma comunque sufficienti per lo scopo dell'analisi.

Per iniziare si osserva come il numero delle recensioni si distribuisce nel corso degli anni e si può subito notare come vi è stato un picco nel 2016 per poi calare e subire una drastica riduzione nel 2020 e 2021, complice delle chiusure e limitazioni apportate dal Covid19 (Figura 3.2).

[33] https://www.tripadvisor.com/TransparencyReport2021

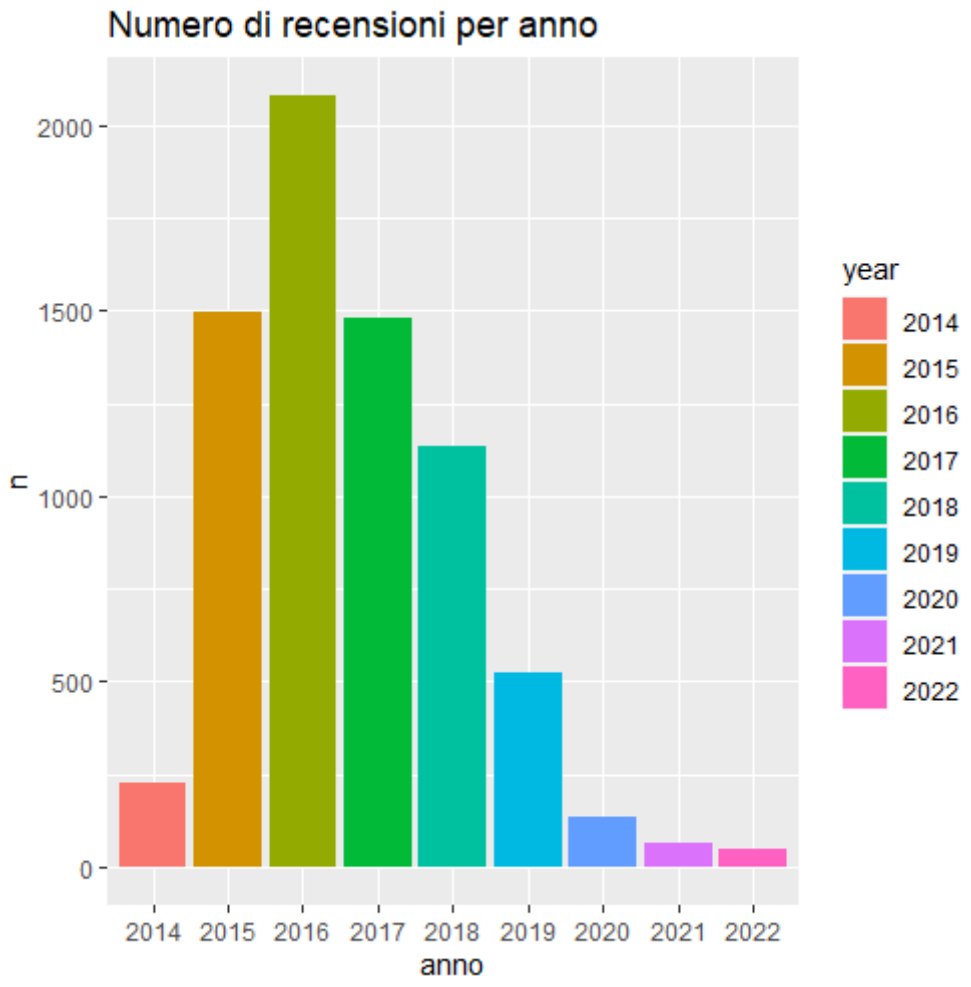


Figura 3.2 Grafico rappresentante le recensioni annue per il MoMA

3.4.1 Analisi delle parole

Il primo passo per eseguire un'analisi su testi è quello di creare un dataset con *tidy data*.

Creare ciò significa impostare la base dati seguendo un determinato standard, ovvero:

- Ogni colonna rappresenta una variabile
- Ogni riga rappresenta un'osservazione
- Ogni cella è un valore singolo

Fatto ciò, possiamo creare una classifica delle parole più utilizzate nelle recensioni per avere così una prima visione generale dei termini presenti.

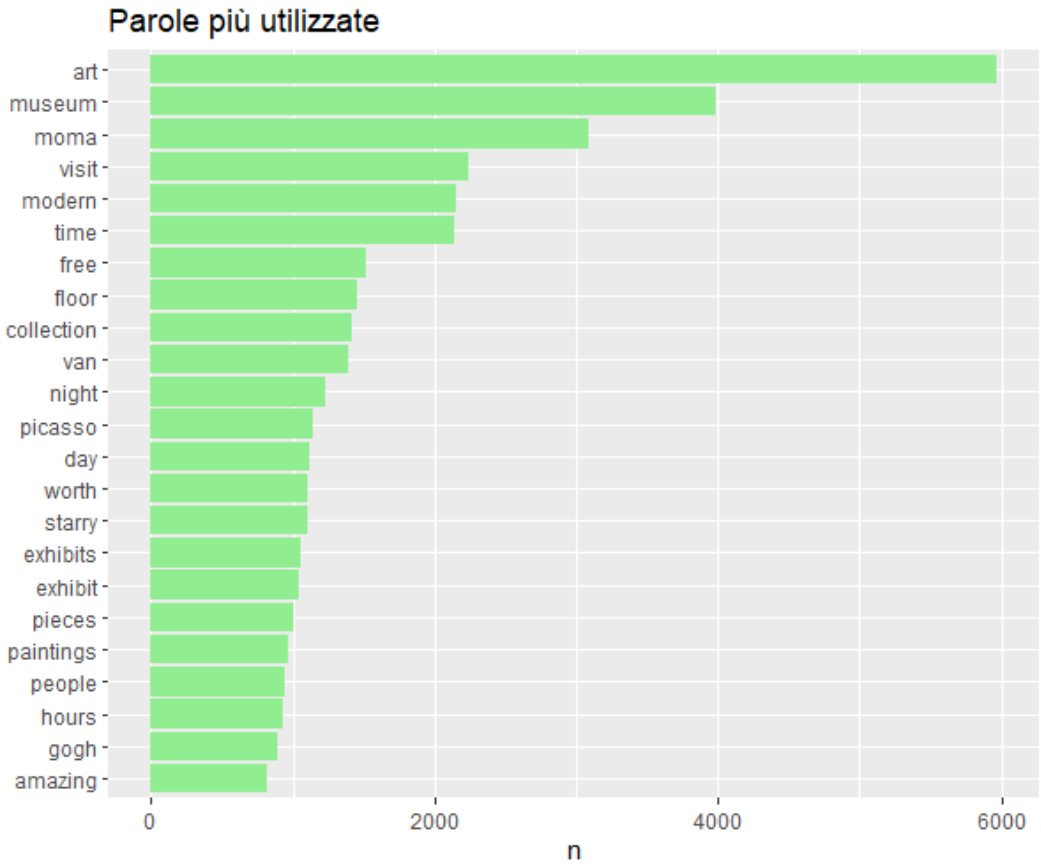


Figura 3.3 Grafico rappresentante le parole più frequenti nelle recensioni

Chiaramente in cima alla classifica troviamo parole generiche per la descrizione dei musei (*museum, art, paintings, collection*, ecc.) o il nome del museo stesso, ma anche descrittori tipici di una visita turistica (*visit, free time, ecc.*). Si può, però, osservare come Van Gogh e Picasso sono gli artisti più citati e l'opera più menzionata sembra essere la "*Starry Night*" di Van Gogh (Figura 3.3, Figura 3.5).

Ripetiamo l'operazione considerando solo la radice delle parole (per esempio *day* e *daily* verranno considerati come *dai-* in un'unica voce) per verificare se si verificano delle differenze.

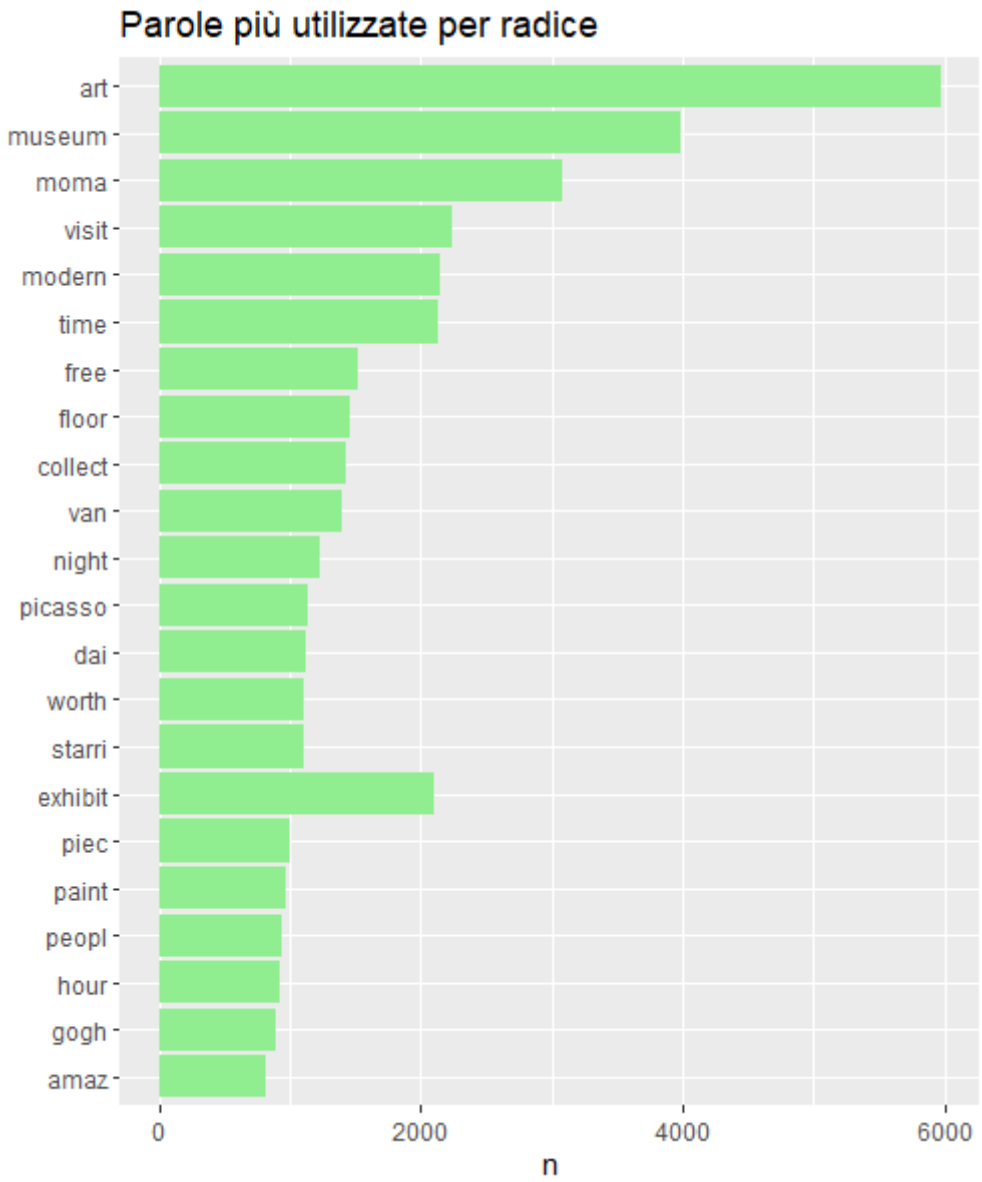


Figura 3.4 Grafico rappresentante le parole più frequenti nelle recensioni con stessa radice

Non sembrano esserci molti cambiamenti se non le parole con radice *exhibit-* e quindi tutto ciò che è legato a *exhibition* che passa ora a oltre 2000 citazioni (Figura 3.4). Un risultato interessante se collegato anche con le informazioni ottenute dalla ricerca sulle immagini di Instagram che poneva la medesima categoria come la più fotografata per il MoMA.

Figura 3.5 Wordcloud rappresentante le parole più utilizzate per le recensioni del MoMA

3.4.2 Sentiment analysis

Si continua effettuando una *sentiment analysis*[34], *o opinion mining*, sul testo a disposizione. Questo tipo di analisi è utilizzato per studiare e quantificare la soggettività dell'informazione. È utilizzata ampiamente per analizzare sondaggi, recensioni e i riscontri dei clienti. Di base, si cerca di comprendere se una parola o una frase ha una valenza positiva, negativa o neutrale. È possibile, inoltre, eseguire un'analisi più avanzata attribuendo lo stato emozionale suscitato: gioia, disgusto, rabbia, sorpresa e altre possibili emozioni. Tecnicamente si tratta di confrontare le parole presenti nel testo considerato con una libreria che, a seconda dello scopo o del metodo, assegna un'emozione o un punteggio ai vocaboli. Ovviamente questa tipologia di analisi condotta con macchine presenta dei limiti causati dall'ambiguità del linguaggio umano, per esempio, fatica con negazioni, esagerazioni, ironia, semplici da cogliere per una persona, ma complesse per un computer. Nonostante le limitazioni, rimane comunque un buon metodo per cogliere le sensazioni generali che il pubblico ha riguardo ad un determinato argomento.

Si inizia analizzando il dataset confrontandolo con la libreria *Bing et al*. che assegna a determinate parole di lingua inglese un valore binario distinto tra positivo e

[34] Feldman, R., 2013. Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), pp.82-89.

negativo in modo da osservare la distribuzione e il sentimento adesso associato delle parole utilizzate (Figura 3.6).

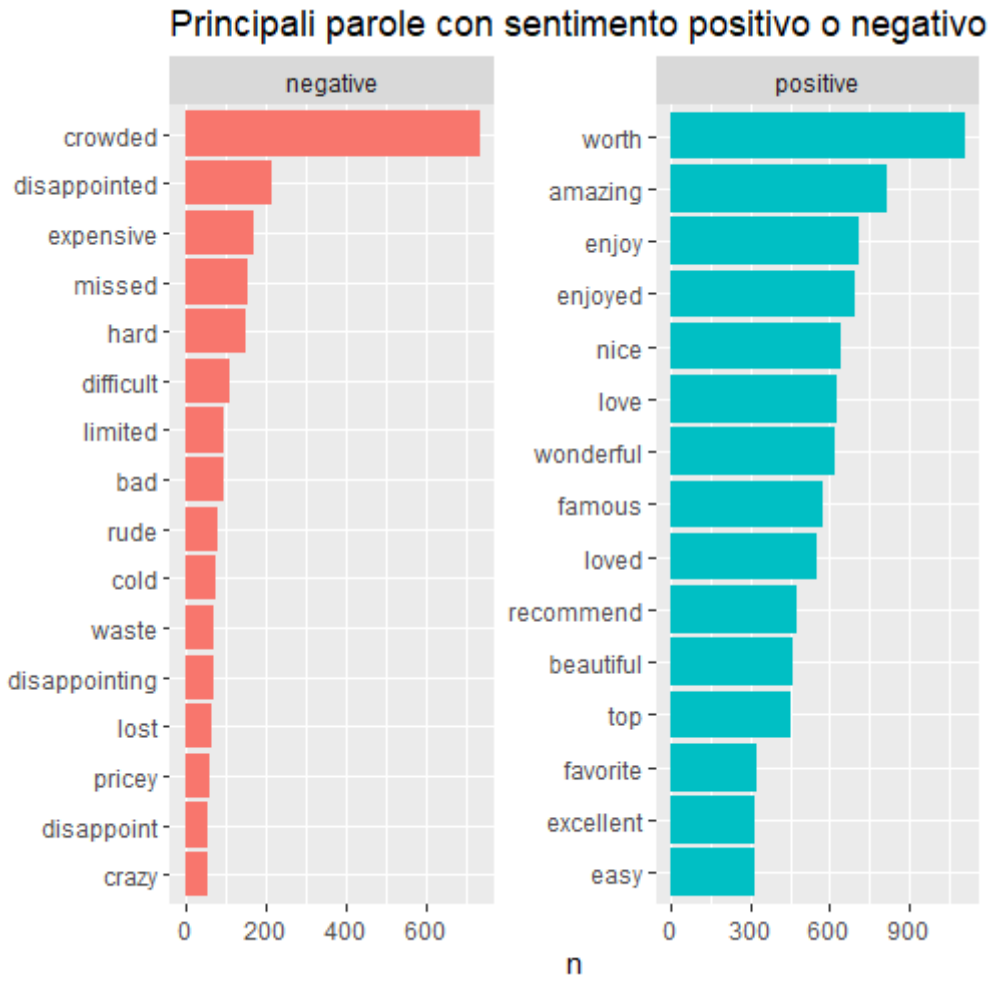


Figura 3.6 Grafico rappresentante le principali parole con sentimento Positivo o Negativo

I risultati ottenuti consentono di osservare che in generale le parole positive sono utilizzate in maggior numero rispetto a quelle negative e come la critica principale sia il sovraffollamento degli spazi ("*crowded*").

Utilizzando la libreria *Affin*, che assegna un punteggio da -5 a +5 a seconda del grado di positività o negatività, si vede come la parola più negativa utilizzata sia deluso ("*disappointed*") (Figura 3.7).

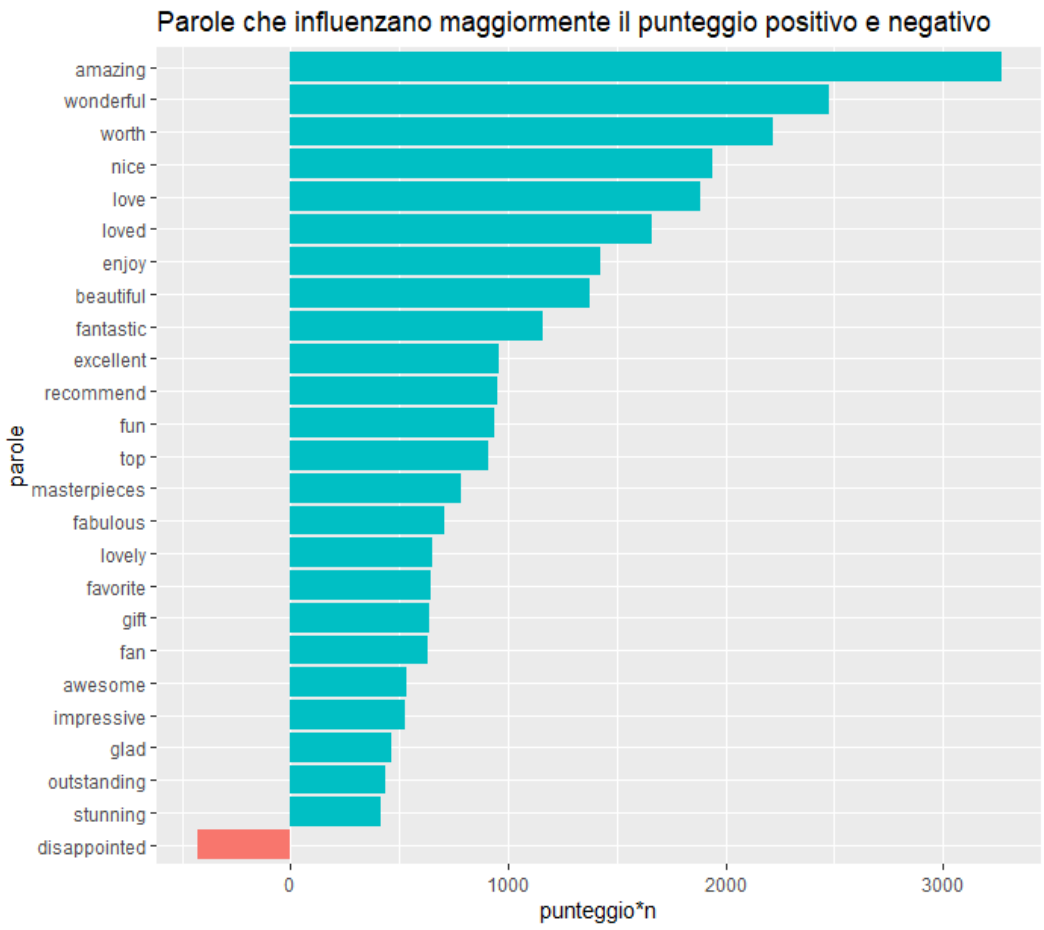


Figura 3.7 Grafico rappresentante le parole che influiscono maggiormente sul punteggio positivo o negativo

Si rischia però di giungere a conclusioni erronee perché le parole in questo caso vengono considerate univocamente e non nel contesto, oppure, semplicemente, non viene considerata la parola che la precede che potrebbe essere una negazione e quindi trasformare qualcosa di positivo in negativo (per esempio *worth* potrebbe essere preceduto da *not*) o viceversa.

Per ovviare al problema si utilizzano i bigrammi, ovvero si considerano le parole a coppie. Dopo avere eseguito questo passaggio possiamo confrontare nuovamente il nostro dataset, utilizzando, nuovamente la libreria *Affin*.

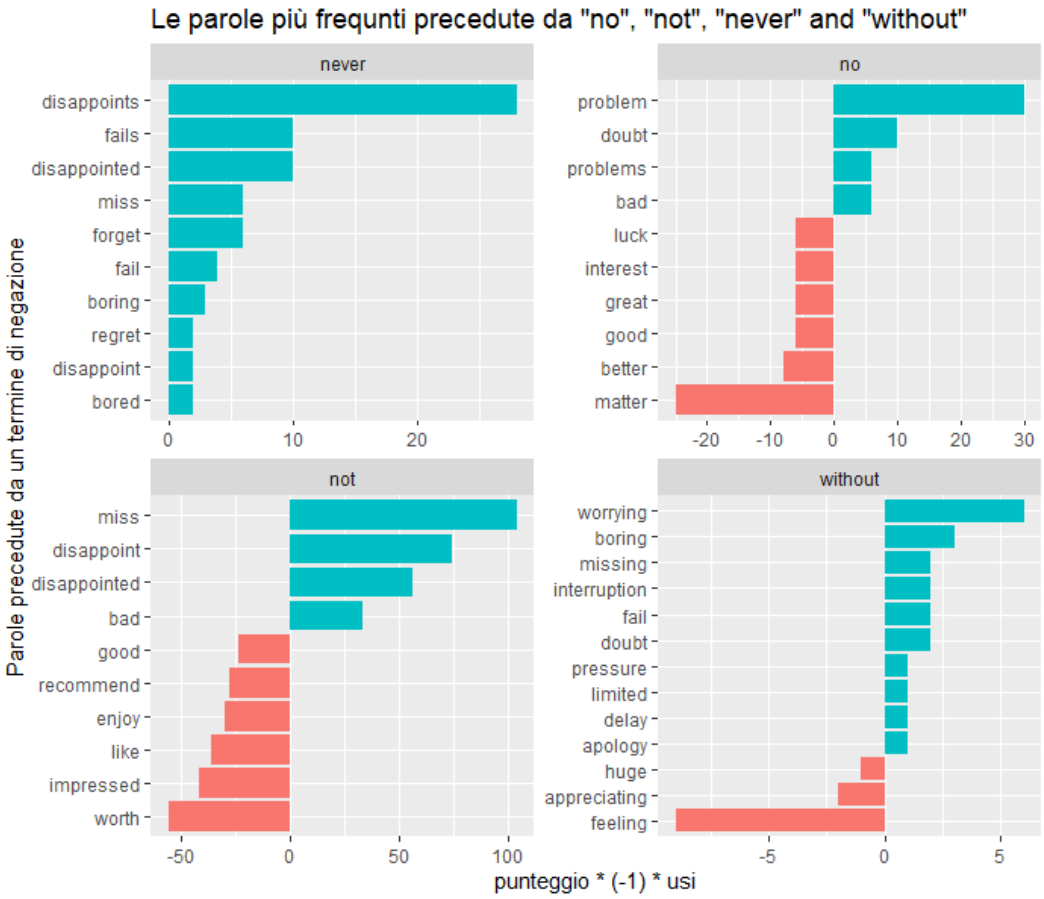


Figura 3.8 Grafico rappresentante le parole precedute da parole di negazione

Sono state considerate le parole più popolari in lingua inglese per introdurre una negazione (*never*, *no*, *not*, *without*) e si nota come la maggiore causa di errore nell'assegnare il giusto sentimento si ha con le parole precedute da *not*, per esempio ora *miss*, quindi *not miss*, aumenta il punteggio positivo di oltre 100 punti, mentre (*not*) *worth* ora lo abbassa di oltre 50 punti (Figura 3.8).

Grazie alla libreria *Affin* è possibile condurre anche un piccolo test e per esempio mostrare quali sono complessivamente le recensioni più positive o negative, sommando per ciascuna di esse il punteggio delle parole usate e facendone poi una media del punteggio ottenuto.

Per esempio, la recensione più positiva risulta essere questa:

"Loved it. The permanent exposition is ok, but in comparison to other museums it is small. HOwever, I loved the temporal expositions. There was one of Charles White which was

beautiful, and breathtaking. There was one of architecture in Yugoslavia which was superb (if you like modernist architecture)."

Mentre quella maggiormente negativa:

"My last visit to MOMA NY was years ago. I went again this week. What an utter disappointment. The logistics are horrible. No clear signs anywhere. Staff is badly informed. The girl who sold me the ticket had no idea there was a book on the 350 most important works of art. The logistics in the galleries on the 4th floor and up are horrid. What MOMA needs is 1. A new director of logistics and 2. A new curator. Contents wise the museum is a mess. MOMA NY is lost....."

3.4.3 Analisi con bigrammi

Con i bigrammi ottenuti in precedenza è possibile visualizzare come tutte le parole si relazionano tra loro (Figura 3.9).

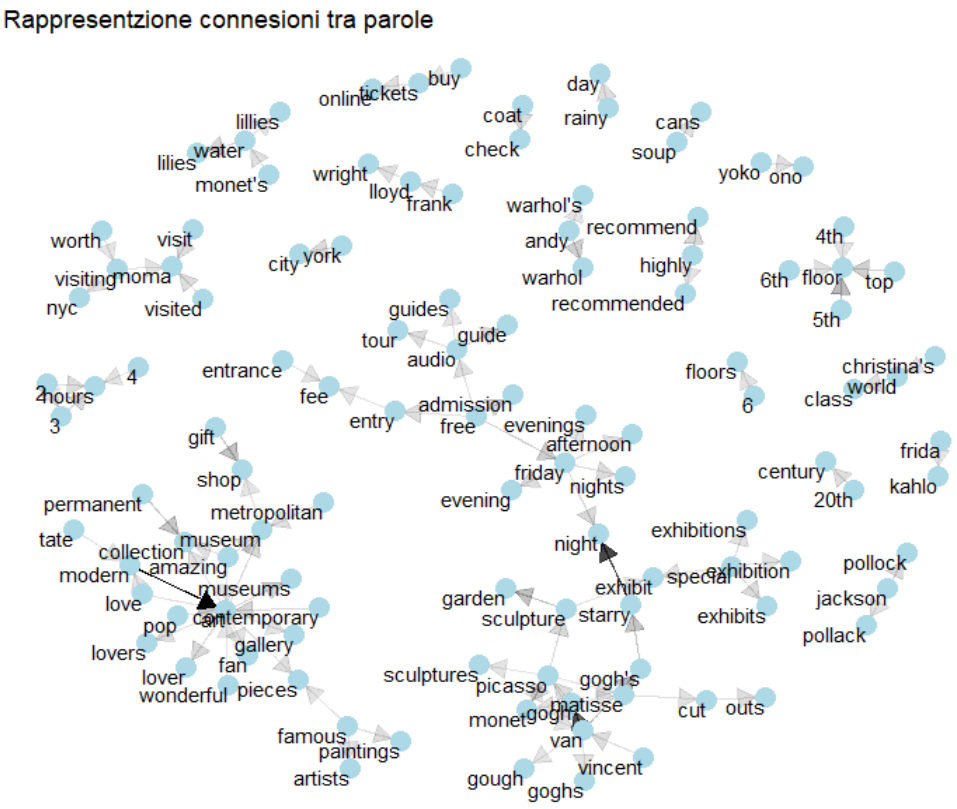


Figura 3.9 Grafico rappresentante la rete di connessione tra le parole utilizzate nelle recensioni

Immediatamente dal grafico si possono trarre molte nozioni interessanti. Si osserva come intuitivamente molte parole sono collegate da un nodo centrale da *art*, mentre possiamo anche ottenere informazioni riguardo ai piani più frequentati dato che il quarto, quinto, sesto e ultimo sono quelli maggiormente menzionati. Inoltre, possiamo anche vedere quali sono gli artisti e le opere più citate: "*Water Lilies*" di Monet, "*Starry Night*" di Van Gogh, "*Soup Cans*" di Andy Warhol, l'esibizione delle sculture nel giardino. Inoltre, si potrebbe dedurre che il giorno preferito dei visitatori sia il venerdì pomeriggio o sera.

3.4.4 Analisi tf-idf

Un'altra tipologia di analisi da considerare è comprendere quale siano le parole più importanti in un documento o, in questo caso, per ogni anno. Per lo scopo si propone di analizzare la frequenza dei termini mediante una combinazione di tecniche utilizzata nell'*information retrival*[35] chiamata tf-idf. Ovvero, si considera la *term frequency* di una parola (tf) e la si moltiplica con la propria inverse *document frequency* (idf) calcolata con la formula seguente[36]:

$$tf * idf(term) = ln\left(\frac{n\ documents}{n\ documents\ containing\ term}\right)$$

Questo permette di quantificare e scoprire quali sono le parole più importanti per un corpo di testo, in questo caso si metteranno in mostra quali sono stati i termini più significativi per ogni anno.

[35] Insieme di tecniche utilizzate per la rappresentazione, la memorizzazione, l'organizzazione e l'accesso ad oggetti contenenti informazioni.
[36] Robertson, S., 2004. Understanding inverse document frequency: on theoretical arguments for IDF. Journal of documentation.

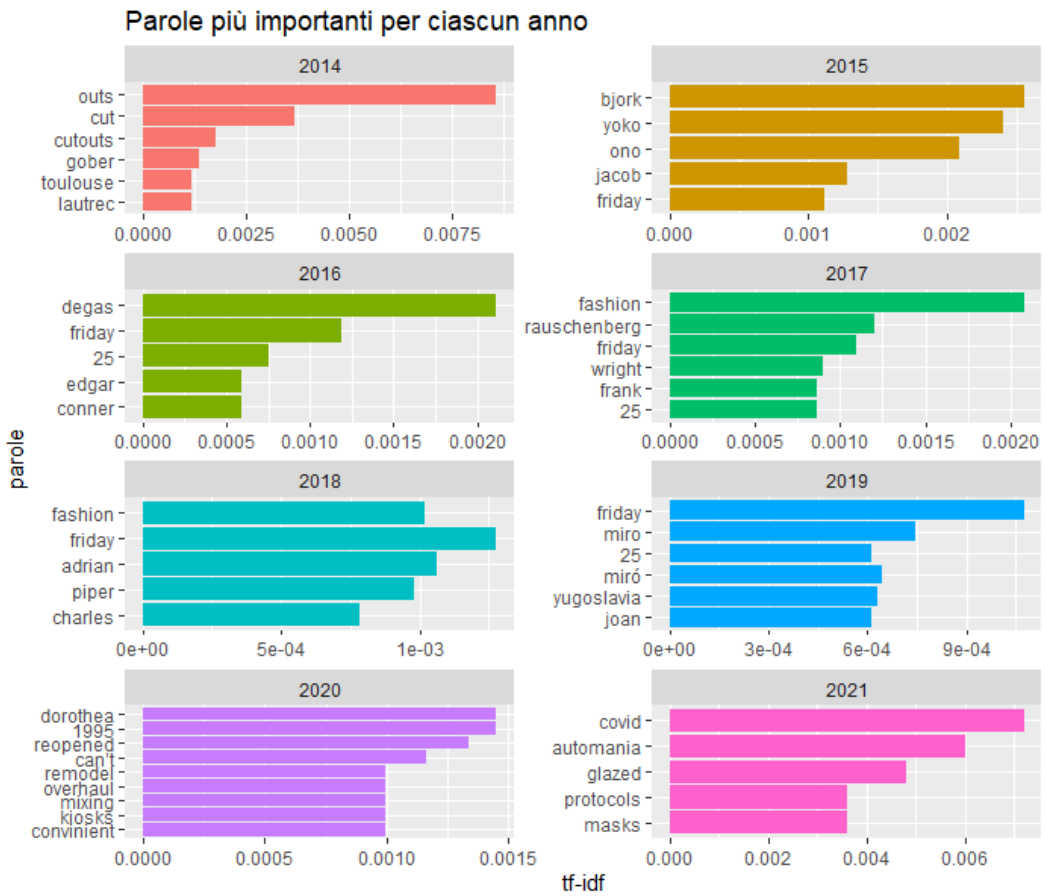


Figura 3.10 Grafico rappresentante le parole più significative per ciascun anno secondo il metodo tf-idf

È possibile notare immediatamente come gli argomenti più importanti per ogni anno riguardano la mostra principale che l'ha caratterizzato.

Come si poteva intuire, le parole più utilizzate nel 2021 sono quelle legate alla pandemia globale Covid19. Infatti, il termine più significativo è stato proprio Covid (Figura 3.10).

Conclusioni

In questa parte finale verranno sintetizzati i risultati ottenuti dalle analisi dei due capitoli precedenti.

L'analisi riguardante gli hashtag di 1800 post relativi ai musei considerati, ha mostrato come per i musei coreani si utilizzano meno hashtag, ma si caricano più foto per ciascun post, i nomi degli artisti e degli oggetti vengono usati raramente, mentre la maggioranza degli hashtag sono di tipo informativo, più precisamente legati al luogo e all'arte in generale. Tutto ciò dimostra il potere attrattivo a livello turistico della località in cui si trovano i musei. Inoltre, nei post sono presenti anche hashtag relativi a luoghi appartenenti ai vari musei, che potrebbero essere utili per comprendere quali sono gli spazi preferiti dei visitatori. Si osserva che non sono presenti hashtag relativi a risposte emotive, quindi, si ritiene che essi non siano uno strumento idoneo a raccogliere informazioni riguardanti le opinioni dei turisti.

Dall'analisi delle immagini si evince una differenza tra i soggetti fotografati a seconda della tipologia di museo considerata. Si osserva che per i musei di arte moderna la categoria *art exhibition* è molto frequente, tra il 25, 64% e il 31,48%, rispetto ai musei generalisti. Le foto della categoria *architecture* e ad *artifact* sono le più frequenti per ogni museo. Mentre le categorie *selfie e human body* sono poco rappresentate per qualsiasi museo, ad eccezione del MoMA dove le classi precedentemente accennate hanno rispettivamente una frequenza del 10,38% e 9,38%.

Questo è dovuto principalmente alla diversa tipologia di oggetti esposti che vanno da quadri, sculture e artefatti classici o storici per i musei generalisti a installazioni di arte moderna per gli altri musei. Grande importanza sembrano avere gli spazi e la struttura del museo, che si ritrovano come elementi piuttosto frequenti nelle foto, indicando come l'organizzazione delle aree adibite all'esposizione e la riconoscibilità dell'edificio giochino un ruolo fondamentale.

Si ritiene che questo tipo di analisi possa essere molto utile per avere una visione generale di ciò che i visitatori ritengono più importante o ciò che attrae maggiormente la

loro attenzione e di conseguenza può rivelarsi un'informazione preziosa per i gestori per selezionare su quali aree fare leva o potenziare.

Il limite di questa ricerca è l'impossibilità di trarre conclusioni sul profilo demografico dei visitatori, che si rivelerebbe essenziale per trarre conclusioni più complete e utili per la scelta di strategie di marketing. Sarebbe opportuno accompagnare il tutto con un'indagine sulle caratteristiche dei turisti in modo da avere una visione più accurata della situazione.

Un'altra mancanza dell'analisi delle immagini è l'impossibilità o la difficoltà a comprendere le reazioni emotive riguardo all'oggetto fotografato o indentificarlo nel dettaglio. Per esempio, questo sistema può riconoscere alcuni elementi iconici come la piramide del Louvre, ma non è in grado di identificare in modo specifico il quadro o l'installazione artistica, né tantomeno comprendere il motivo per cui è stato condiviso.

Si reputa interessante, di conseguenza, prestare attenzione alle opinioni dei visitatori dalle quali si possono trarre informazioni sulla loro risposta emotiva, per questo un'indagine sulle recensioni potrebbe essere idonea a raccogliere importanti indicazioni sugli effettivi pareri dei visitatori.

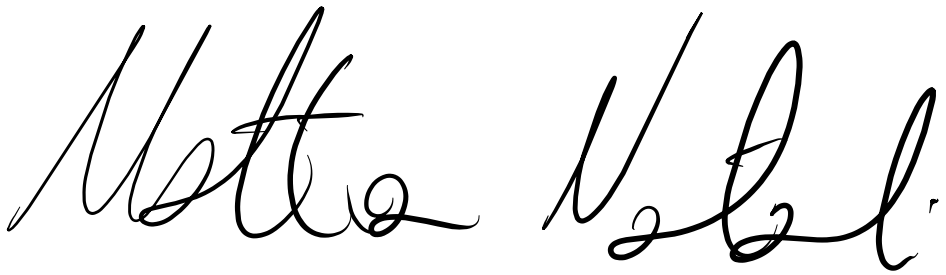
L'indagine sulle recensioni si è posta come obiettivo quello di ottenere una visione generale delle opinioni dei visitatori in modo da poter aggiungere informazioni ulteriori da affiancare alla sintesi rappresentata da un punteggio. Mediante una *sentiment analysis* si è potuto osservare che le recensioni hanno per la maggior parte contenuti positivi, mentre le critiche mosse con più frequenza riguardano l'affollamento e una visita deludente. L'analisi dei bigrammi ha permesso di ottenere informazioni relative ai piani più discussi: il quarto, il quinto e il sesto, le ore di permanenza: tra le 2 e le 4 ore e gli artisti più menzionati assieme alle loro opere: per esempio Van Gogh e il quadro "*Starry Night"*. La statistica tf-idf invece ha mostrato gli argomenti più importanti per ciascun anno considerato, generalmente il topic più discusso riguardava una mostra, fa eccezione il 2021 dove i termini riguardanti la pandemia sono risultati i più significativi.

Ovviamente l'intera operazione presenta grossi limiti legati alla complessità del linguaggio, per esempio l'incapacità di comprendere l'ironia o significati ambigui delle parole e frasi. Numerosi lavori sono in corso per ampliare e perfezionare tecniche per una

sempre migliore comprensione e analisi del linguaggio umano, specialmente si stanno facendo progressi grazie all'utilizzo dell'intelligenza artificiale.

In conclusione, sono stati mostrati alcuni esempi di facile e rapida implementazione che sfruttano strumenti alla portata di tutti: l'analisi per immagini e delle recensioni partono da social media o piattaforme gratuite (Instagram e TripAdvisor), nonché sono stati utilizzati strumenti e sistemi fruibili gratuitamente o con piani premium opzionali permettendo a tutti i musei dai più piccoli ai più grandi, da quelli con risorse limitate a quelli con ampli budget, di osservare e comprendere qualcosa in più riguardo ai propri visitatori. Grazie a questi metodi si è potuto raccogliere in modo veloce e semplice dati relativi a ciò che gli utenti condividono più frequentemente sui social quando visitano un museo, dando quindi informazioni ai gestori su quale aree puntare o rafforzarsi. Inoltre, è possibile analizzando le recensioni ricavare la risposta emotiva del pubblico nei confronti dell'istituzione utile per lavorare sulla propria immagine e brand e informazioni relative agli argomenti, opere e temi più discussi dai turisti.

Il grosso limite di tutto ciò è rappresentato dalla quantità di dati necessaria per ottenere risultati attendibili, ovviamente musei di grande rilevanza internazionale come il British Museum o MoMA possono contare su decine di migliaia di post e recensioni su cui lavorare, mentre i piccoli musei hanno numeri molto più contenuti. Per questo motivo per questa categoria di musei potrebbe essere più utile sfruttare questo tipo di analisi per studiare la concorrenza ed eventualmente mimare le strategie adottate da essi quando possibile oppure operare con tecnologie applicabili localmente: tracciamento degli spostamenti all'interno del museo, strumenti offerti dai social media, scanner per la risposta emotiva dei visitatori posti vicino alle opere e altra apparecchiatura o software utile per ottenere feedback su ciò che coglie l'interesse all'interno del museo.



Bibliografia

A. Gagliardi, Uffizi al top per visitatori in Italia, superato il Colosseo: pesa anche l'effetto Ferragni, Il Sole 24 Ore, 5 aprile 2022

Alessandrini, S., 2015. Un modello di simulazione dell'impatto economico dei musei/istituti dell'area POIN: metodo, fonte dati, primi risultati.

Ayeh, J.K., Au, N. and Law, R., 2013. "Do we believe in TripAdvisor?" Examining credibility perceptions and online travelers' attitude toward using user-generated content. Journal of Travel Research, 52(4), pp.437-452.

Baggio, R., 2010. Recensioni online: una rassegna della ricerca scientifica. BTO Educational Occasional Paper, Milano.

Brown Jarreau, P., Dahmen, N.S. and Jones, E., 2019. Instagram and the science museum: A missed opportunity for public engagement. Journal of Science Communication, 18(2), p.A06.

Budge, K. 2017. Objects in focus: Museum visitors and Instagram. Curator: The Museum Journal 60(1): 67-85.

Budge, K. 2018. Encountering people and place: Museums through the lens of Instagram. Australasian Journal of Popular Culture 7(1): 107-121.

Budge, K. and Burness, A., 2018. Museum objects and Instagram: agency and communication in digital engagement. Continuum, 32(2), pp.137-150.

Budge, K., 2017. Objects in focus: Museum visitors and Instagram. Curator: The Museum Journal, 60(1), pp.67-85.

Carah, N., 2014. Curators of databases: Circulating images, managing attention and making value on social media. Media International Australia, 150(1), pp.137-142.

Chlebus-Grudzień, P., 2018. Selfie at a museum: Defining a paradigm for an analysis of taking (self-portrait) photographs at museum exhibitions. Turyzm, 28(1), pp.7-13.

CMA, 2015, Online reviews and endorsement: Report on the CMA's call for information

Feinerer, I., Hornik, K. and Meyer, D., 2008. Text mining infrastructure in R. Journal of statistical software, 25, pp.1-54.

Feldman, R., 2013. Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), pp.82-89.

Giaccardi, E. and Palen, L. 2008. The Social Production of Heritage through Cross-media Interaction: Making Place for Place-making. International Journal of Heritage Studies 14(3): 281-297.

Holdgaard, N. and Klastrup, L., 2014. Between control and creativity: challenging co-creation and social media use in a museum context. Digital Creativity, 25(3), pp.190-202.

Ibba, S., Orrù, M., Pani, F.E. and Porru, S., 2015, November. Hashtag of Instagram: From Folksonomy to Complex Network. In KEOD (pp. 279-284).

ICOM, 2021, Museums, museum professionals and Covid-19: third survey

Jarreau, B., Dahmen P., Smith N., and Jones, E. 2019. Instagram and the science museum: a missed opportunity for public engagement. Journal of Science Communication 18(2): 1-22.

Kidd, J., 2011. Enacting engagement online: framing social media use for the museum. Information Technology & People.

Mikolov, T., Chen, K., Corrado, G. and Dean, J., 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Oh, C., Lee, T., Kim, Y., Park, S., and Suh, B. 2016. Understanding Participatory Hashtag Practices on Instagram: A Case Study of Weekend Hashtag Project. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16): 1280–1287.

Ordenes, F.V. and Zhang, S., 2019. From words to pixels: text and image mining methods for service research. Journal of Service Management.

Pittman, M. and Reich, B., 2016. Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. Computers in Human Behavior, 62, pp.155-167.

Rauschnabel, P. A., Sheldon, P., and Herzfeldt, E. 2019. What motivates users to hashtag on social media? Psychology & Marketing 36(5): 473-488.

Rhee, B. A., Pianzola, F., and Choi, G. T. (Mar. 10, 2020). Collection of categorized Google Vision API image tags for museums related pictures. Zenodo. http://doi.org/10.5281/zenodo.3706462

Rhee, B. A., Pianzola, F., and Choi, G. T. 2020 Analyzing the museum experience through the lens of Instagram posts

Robertson, S., 2004. Understanding inverse document frequency: on theoretical arguments for IDF. Journal of documentation.

Russo, A., Watkins, J. and Groundwater-Smith, S., 2009. The impact of social media on informal learning in museums. Educational media international, 46(2), pp.153-166.

Silge, J. and Robinson, D., 2016. tidytext: Text mining and analysis using tidy data principles in R. Journal of Open-Source Software, 1(3), p.37.

Silge, J. and Robinson, D., 2022. Welcome to text mining in R.

Su, L., Yang, Q., Swanson, S.R. and Chen, N.C., 2021. The impact of online reviews on destination trust and travel intention: The moderating role of online review trustworthiness. Journal of Vacation Marketing, p.13567667211063207.

Suess, A. 2018. Instagram and Art Gallery Visitors: Aesthetic experience, space, sharing and implications for educators. Australian art education 39(1): 107-122.

Usai, A., Pironti, M., Mital, M. and Mejri, C.A., 2018. Knowledge discovery out of text data: a systematic review via text mining. Journal of Knowledge Management.

Yoshimura, Y., Sobolevsky, S., Ratti, C., Girardin, F., Carrascal, J.P., Blat, J. and Sinatra, R., 2014. An analysis of visitors' behavior in the Louvre Museum: A study

using Bluetooth data. Environment and Planning B: Planning and Design, 41(6), pp.1113-1131.

Zingone, M. 2019. Instagram as Digital Communication Tool for the Museums: a Reflection on Prospectives and Opportunities through the Analysis of the Profiles of Louvre Museum and Metropolitan Museum of New York. European Journal of Social Science Education and Research 6(3): 53-63.

Sitografia

- ICOM Italia, definizione musei

  https://www.icom-italia.org/definizione-di-museo-di-icom/

- Art-Museum, dati e articoli su musei italiani

  https://www.art-museum.it/social-media-marketing-per-museo.html

- Progetto ShareArt Bologna Musei

  https://www.enea.it/it/Stampa/news/beni-culturali-al-via-il-progetto-shareart-per-misurare-il-gradimento-delle-opere-d2019arte-con-l2019intelligenza-artificiale

- Instagram sorpassa i due miliardi di utenti

  https://www.cnbc.com/2021/12/14/instagram-surpasses-2-billion-monthly-users.html

- Sondaggio consumatori Brightlocal

  https://www.brightlocal.com/research/local-consumer-review-survey/

- Transparency report TripAdvisor

  https://www.tripadvisor.com/TransparencyReport2021

- Indagine di Which? su presunte recensioni false su TripAdvisor

  https://www.which.co.uk/news/article/fake-tripadvisor-reviews-push-worlds-best-hotels-up-the-rankings-ayEVG8p3qiXY

- Articolo di Brightlocal su recensioni false su Amazon

  https://www.brightlocal.com/learn/review-management/fake-reviews/why-are-fake-reviews-a-problem/#

- Similarweb informazioni sulla popolarità dei siti web

  https://www.similarweb.com/it/top-websites/

- Libro Text Mining in R open source

  https://www.tidytextmining.com/

- Articolo di Susan Li sull’analisi delle recensioni di TripAdvisor per un Hotel

  https://towardsdatascience.com/scraping-tripadvisor-text-mining-and-sentiment-analysis-for-hotel-reviews-cc4e20aef333

- Google Vision Api

  https://cloud.google.com/vision

- Cran

  https://cran.r-project.org/

Indice delle figure

Figura 1.1 Esempio interfaccia dell'app Geronimo Stilton Adventures 6

Figura 1.2 Tweet postato da The MERL diventato virale 10

Figura 1.3 Esempio Tweet postato da Tate Modern 11

Figura 1.4 Esempio dei dati disponibili tramite Instagram 13

Figura 2.1 Attribuzione tag di Google Vision 20

Figura 2.2 Esempi di suddivisione categorie Art exhibition (a), Artifact (b), Architecture (c), Food (d), Human body (e), Landscape (f), manca l'esempio di un selfie per motivi di privacy. 22

Figura 2.3 Grafico rappresentante le etichette più frequenti per i post legati al Louvre 23

Figura 2.4 Wordcloud rappresentante le etichette più frequenti per i post legati al Louvre 23

Figura 2.5 Grafico rappresentante le etichette più frequenti per i post legati al Pompidou 24

Figura 2.6 Wordcloud rappresentante le etichette più frequenti per i post legati al Pompidou 24

Figura 2.7 Grafico rappresentante le etichette più frequenti per i post legati al Met 25

Figura 2.8 Wordcloud rappresentante le etichette più frequenti per i post legati al Met 25

Figura 2.9 Grafico rappresentante le etichette più frequenti per i post legati al MoMA 26

Figura 2.10 Wordcloud rappresentante le etichette più frequenti per i post legati al MoMA 26

Figura 2.11 Grafico rappresentante le etichette più frequenti per i post legati al BM 27

Figura 2.12 Wordcloud rappresentante le etichette più frequenti per i post legati al BM 27

Figura 2.13 Grafico rappresentante le etichette più frequenti per i post legati al Tate 28

Figura 2.14 Wordcloud rappresentante le etichette più frequenti per i post legati al Louvre 28

Figura 2.15 Grafico rappresentante la distribuzione delle categorie per ciascun museo 30

Figura 3.1 Esempio Recensione di TripAdvisor 33

Figura 3.2 Grafico rappresentante le recensioni annue per il MoMA 36

Figura 3.3 Grafico rappresentante le parole più frequenti nelle recensioni 37

Figura 3.4 Grafico rappresentante le parole più frequenti nelle recensioni con stessa radice 38

Figura 3.5 Wordcloud rappresentante le parole più utilizzate per le recensioni del MoMA 39

Figura 3.6 Grafico rappresentante le principali parole con sentimento Positivo o Negativo 40

Figura 3.7 Grafico rappresentante le parole che influiscono maggiormente sul punteggio positivo o negativo 41

Figura 3.8 Grafico rappresentante le parole precedute da parole di negazione 42

Figura 3.9 Grafico rappresentante la rete di connessione tra le parole utilizzate nelle recensioni 44

Figura 3.10 Grafico rappresentante le parole più significative per ciascun anno secondo il metodo tf-idf 45

Indice delle tabelle

Tabella 2.1 Numero visitatori e followers musei considerati 18

Tabella 2.2 Analisi quantitativa hashtag per i post 19

Tabella 2.3 Riassunto categorizzazione hashtag 19

Tabella 2.4 Riassunto statistiche di attribuzione etichette di Google Vision 21