

Valutazione e confronto dell'attrattività turistica tra le principali città europee

Università degli Studi di Milano Bicocca,
Dipartimento di Informatica, Sistemistica e Comunicazione (DISCO),
Facoltà di Data Science,
Anno Accademico 2019/2020

Davide Garavaldi 818308, Davide Porcellini 816586, Simone Tufano 816984

Nel seguente articolo vengono riportate le tecniche utilizzate e le analisi svolte nella realizzazione del progetto del corso di Data Management and Data Visualization. Inizialmente, verrà mostrato il quesito che è stato preso in esame e verranno descritte le fonti utilizzate per la raccolta dei dati. Successivamente, si mostreranno i procedimenti effettuati negli script per trasformare i dati e quindi lo schema integrato. Infine, si concluderà illustrando le visualizzazioni ottenute, oltre ai punti di criticità e di forza dello studio. Il progetto è stato svolto in parte equa da tutti i componenti del gruppo.

1. Introduzione e Scopo

In un mondo sempre più globalizzato come quello in cui viviamo oggi, viaggiare non è mai stato così semplice, soprattutto all'interno della cornice europea: infatti, tra voli low-cost e svariati siti internet che forniscono consigli riguardo a vitto e alloggio nelle diverse città (primo fra tutti Tripadvisor), il viaggio è diventato quasi un "hobby", un passatempo amato da tutte quelle persone che fanno della curiosità di conoscere posti nuovi un'interessante attività. È proprio da questa concezione di viaggio che nasce la domanda iniziale di questo progetto: *"Vacanza in città: dove andare?"*.

L'obiettivo dello studio, infatti, è stato quello di supportare un potenziale turista nella scelta di una meta per una futura vacanza. A questo proposito sono state confrontate le principali città europee sotto vari aspetti, dalla cultura ai trasporti, dal vitto all'alloggio, considerati sia per ciò che le città offrono sia per i relativi costi, facendo riferimento a dati inerenti all'anno 2019.

2. Descrizione delle fonti

Lo studio analizza le dimensioni di Varietà e Velocità: la prima è stata ottenuta dalle diverse fonti prese in esame e che verranno approfondite in seguito, mentre la seconda è stata ricavata grazie a dati "ascoltati" dal social network Twitter.

2.1 Eurostat

Il primo riferimento per reperire dati relativi alle città europee è stato il sito Eurostat. Nello specifico, sono state scaricate le serie storiche di diversi indicatori relativi alle seguenti macrocategorie: Ambiente, Condizioni di Vita, Cultura e Turismo, Popolazione, Trasporti.

Per ogni macrocategoria sono stati selezionati gli indicatori più opportuni per la costruzione di una futura classifica e più importanti per la scelta di una meta turistica. In particolare:

- **Ambiente:** Rainfall - litre/m², Share of land (%): Green urban areas and sports and leisure facilities, Share of land (%): Transport infrastructure, Average temperature of coldest month, Average temperature of warmest month;
- **Condizioni di Vita:** Number of murders and violent deaths;
- **Cultura e Turismo:** Total nights spent in tourist accommodation establishments, Total nights spent in tourist accommodation establishments per resident population, Number of available beds per 1000 residents, Number of bed-places in tourist accommodation establishments, Number of museum visitors (per year), Number of public libraries (all distribution points);
- **Popolazione:** Total population on the 1st of January;
- **Trasporti:** Length of bicycle network (dedicated cycle paths and lanes) - km, People killed in road accidents per 10000 pop. Share of journeys to work by bicycle -%, Share of journeys to work by car or motor cycle -%, Share of journeys to work by foot -%, Share of journeys to work by public transport (rail, metro, bus, tram) -%, Cost of a combined monthly ticket (all modes of public transport) for 5-10 km in the central zone - EUR, Cost of a taxi ride of 5 km to the centre at day time - EUR;

I dataset relativi ad ogni indice sono composti da oltre 1000 osservazioni corrispondenti alle città e 29 colonne corrispondenti agli anni delle serie storiche dal 1990 al 2018.

Sono stati eliminati i duplicati ed è stata effettuata una pulizia dei dati, poiché i valori delle variabili erano spesso accompagnati da caratteri alfabetici che ne rendevano impossibile l'utilizzo; i nomi delle città scritti in lingua originale sono stati tradotti in lingua inglese.

Dato che l'analisi si basa sul confronto di dati relativi al 2019, sono stati previsti, dove mancanti, i valori degli indici di tale anno tramite una regressione lineare applicata ad ogni osservazione, in modo da ottenere dati omogenei per l'anno considerato. Le città che presentavano solo dati mancanti sono state rimosse; per le città con dati parzialmente mancanti, questi stessi dati sono stati imputati con la media di riga.

Calcolati i valori previsti per tutti gli indicatori, si è deciso di creare un unico dataset integrato, contenente tutti gli indici presi dal sito Eurostat e di caricarlo su Mongo.

```
{
  "_id": ObjectID("5e1b057c41b104b9ff8d76e8"),
  "Share of journeys to work by public transport (rail, metro, bus, tram)": 675.1371551724138,
  "city": "Bruxelles / Brussel",
  "Cost of a combined monthly ticket (all modes of public transport) for 5-10 km in the central zone": 35.96147619047622,
  "Number of public libraries": 31.913669950738914,
  "Average temperature of warmest month": 19.329885057471266,
  "Green urban areas and sports and leisure facilities": 13.199999999999996,
  "People killed in road accidents per 10000 pop": 2.0760098522167483,
  "Total population on the 1st of January": 1199624.6403940879,
  "Average temperature of coldest month": 2.2503284072249587,
  "Total nights spent in tourist accommodation establishments per resident population": 27.825533661740558,
  "Share of journeys to work by car or motor cycle": 51.406317733990136,
  "Length of bicycle network ": 41.52229064039408,
  "Transport infrastructure": 41.7,
  "Share of journeys to work by bicycle": 4.892044334975369,
  "Number of available beds per 1000 residents": 36839.657224958966,
  "Rainfall": 824.188095238095,
  "Number of murders and violent deaths": 35.396934865900334,
  "Number of bed-places in tourist accommodation establishments": 5772217.351395741,
  "Share of journeys to work by foot": 0.2283060755336619,
  "Cost of a taxi ride of 5 km to the centre at day time": 11.965250410509029,
  "Number of museum visitors": 5.026683087027919,
  "Total nights spent in tourist accommodation establishments": 1177071.408374384
}
```

2.2 Kaggle

Da Kaggle sono stati scaricati 3 dataset relativi a recensioni di Tripadvisor riguardanti alberghi, ristoranti e musei.

Il primo dataset analizzato è stato relativo ai musei. È stato creato un dataset ridotto con le seguenti variabili di interesse: Address, Rating, ReviewCount, free entry, State, 'Latitude', 'Longitude'.

Tale dataset è stato filtrato per ottenere i musei relativi a Stati europei; poiché erano presenti indirizzi riferiti ad aree periferiche, si è deciso di utilizzare le coordinate geografiche per includere nelle città, ritenute più interessanti per lo studio, i musei presenti nel raggio di 25 km (successivamente si utilizzerà la stessa misura per georeferenziare i tweet).














Sono state create le variabili necessarie per calcolare il rating medio, pesato in base al numero delle recensioni ed è stato ottenuto per le 16 città trovate. Il documento è stato caricato su Mongo.

Successivamente, è stato analizzato il file relativo agli hotel, i cui dati si riferivano solo a 6 città. Tale file è stato caricato ugualmente su Mongo per valutare successivamente la possibilità di ottenere dati relativi agli hotel per altre città.

Infine, sono stati analizzati i ristoranti con gli stessi procedimenti, ottenendo dati relativi a 31 città.

```
{
  "_id": ObjectID("5e1b2c500a2574304ccb4c23"),
  "city": "Amsterdam",
  "museum_rating": 4.5,
  "restaurant_rating": 4.154251237802936,
  "hotel_rating": 8.362745063477444
}
```

2.3 Numbeo

Do you live in Paris ? Add data for Paris!		
Currency: EUR Sticky Currency Switch to US measurement units		
Restaurants	[Edit]	Range
Meal, Inexpensive Restaurant	15.00 €	10.00  18.00
Meal for 2 People, Mid-range Restaurant, Three-course	50.00 €	38.00  65.00
McMeal at McDonalds (or Equivalent Combo Meal)	8.50 €	8.00  10.00
Domestic Beer (0.5 liter draught)	7.00 €	5.00  8.00
Imported Beer (0.33 liter bottle)	6.00 €	4.00  7.00
Cappuccino (regular)	3.34 €	2.00  4.80
Coke/Pepsi (0.33 liter bottle)	3.11 €	2.00  4.00
Water (0.33 liter bottle)	2.16 €	1.00  4.00
Markets	[Edit]	
Milk (regular), (1 liter)	1.23 €	0.90  1.70
Loaf of Fresh White Bread (500g)	1.51 €	1.00  2.50
Rice (white), (1kg)	1.75 €	1.00  3.00
Eggs (regular) (12)	3.32 €	2.28  4.80
Local Cheese (1kg)	18.13 €	10.00  30.00

scaricati dal sito in versione italiana ed è stato selezionato l'euro come valuta.

Una volta ottenuto il dataset completo, le città e gli indicatori sono stati tradotti in inglese per uniformare i dati a quelli ottenuti da Eurostat e sono stati caricati su Mongo.

Numbeo è un sito web che riporta per ogni città il costo della vita, analizzando molteplici voci (es. vitto, alloggio, trasporti, ecc...). Si è deciso, quindi, di creare una lista formata dalle 31 città ottenute dal file dei ristoranti e fare “scraping” per ottenerne i relativi dati. Questi sono stati

2.4 Twitter

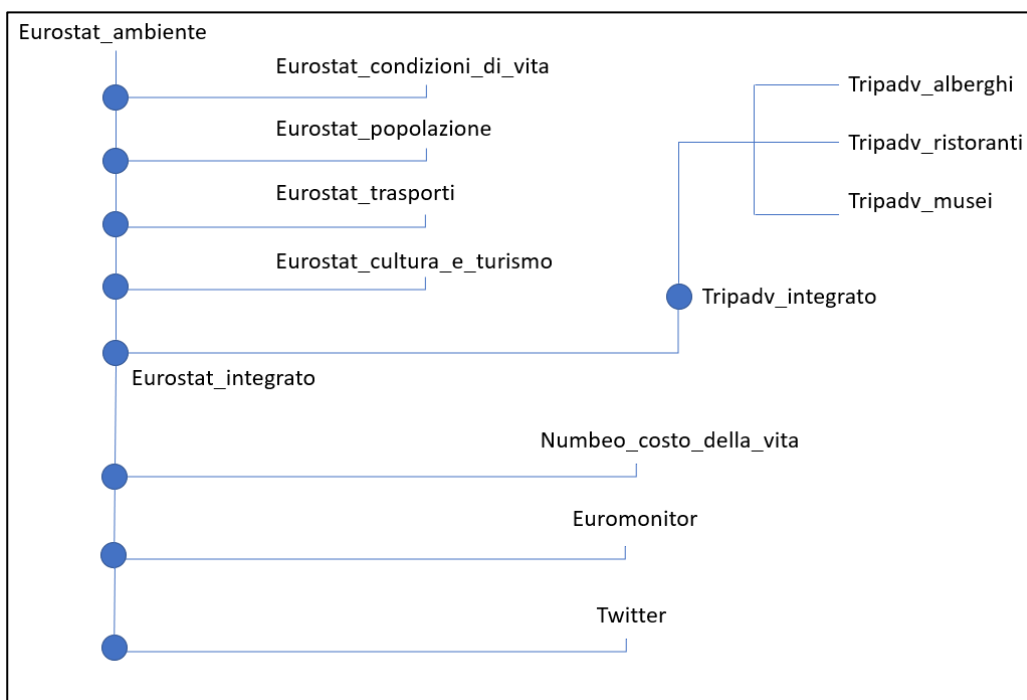
Per valutare la popolarità delle città si è deciso di calcolare il numero di tweet georeferenziati nel raggio di 25 km in cui la città fosse menzionata. La velocità è stata simulata acquisendo i tweet mese per mese per l'anno 2019 e inviandoli regolarmente al Consumer grazie al software Kafka. È stato calcolato il totale dei tweet per ogni città e il file ottenuto è stato caricato su Mongo.

```
{
  "_id": ObjectId("5e1c741b8bb44b1c7b8b66e8"),
  "city": "Barcelona",
  "jan": 9643,
  "feb": 9762,
  "mar": 14012,
  "apr": 14622,
  "may": 15955,
  "jun": 10107,
  "jul": 3100,
  "aug": 10916,
  "sep": 11768,
  "oct": 16596,
  "nov": 13770,
  "dec": 12692,
  "total": 142943
}
```

2.5 Euromonitor

Euromonitor è il principale fornitore indipendente al mondo di ricerche strategiche di mercato. Da questo sito web sono stati scaricati i dati relativi alle 100 città del mondo che vantano il maggior numero di turisti, registrato nel 2019. Successivamente, sono state ricercate e selezionate le città presenti anche nella lista ottenuta dal file dei ristoranti. Da tale incrocio sono state ricavate 25 città. Infine, il file è stato caricato su Mongo.

3. Schema di arricchimento e integrazione



Nello schema sopra, viene illustrato lo schema di arricchimento, con le differenti collezioni create. Successivamente, tutti i documenti sono stati integrati in unico dataset, che è stato memorizzato in un database non relazionale (Mongodb) in formato json. Di seguito ne viene mostrata la struttura:

```
{
  '_id':..., 'city':..., 'Share of journeys to work by public transport (rail,
metro, bus, tram)':..., 'Share of journeys to work by car or motor cycle':...,
ecc..., 'Meal in Inexpensive Restaurant':..., ecc.., 'Number of bed-places in
tourist accommodation establishments':..., 'total tweets':..., 'Number of public
libraries':..., ecc..
}
```

I motivi della scelta di questo formato sono principalmente 3:

- 1) Parte dei dati di partenza erano in formato json;
- 2) La flessibilità dello schema di questo formato permette di salvare gli indicatori necessari per le città considerate senza preoccuparsi del fatto che alcune città presentino dati mancanti su alcune variabili;
- 3) In futuro, per analizzare le variazioni dell'attrattività turistica, potranno essere aggiunte città che presentano indicatori differenti e valutarli secondo le rispettive serie storiche, tralasciando il problema della modifica dello schema.

4. Creazione degli indicatori

Una volta ottenuto il database finale, si è pensato di creare degli indicatori generali che descrivessero le città sotto 7 punti di vista: Trasporti, Vitto, Alloggio, Cultura, Popolarità, Tempo libero e Sicurezza. Dal momento che numerose città presentavano valori mancanti in diversi aspetti, si è deciso di restringere l'analisi alle sole 27 città che non presentassero valori mancanti per alcun aspetto. Di seguito si riportano i singoli indici utilizzati per creare gli indicatori generali:

- **Trasporti:** Share of journeys to work by public transport (rail, metro, bus, tram), Share of journeys to work by car or motor cycle, Length of bicycle network, Transport infrastructure, Share of journeys to work by bicycle, Share of journeys to work by foot, One-way Ticket (Local Transport), Monthly Pass (Regular Price), Taxi Start (Normal Tariff), Gasoline (1 liter);
- **Vitto:** Meal in Inexpensive Restaurant, Meal for 2 People Mid-range Restaurant Three-course, McMeal at McDonalds (or Equivalent Combo Meal), Domestic Beer (0,5 liter draught) Restaurant, Imported Beer (0,33 liter bottle) Restaurant, Cappuccino (regular), Coke/Pepsi (0,33 liter bottle), Water (0,33 liter bottle), Milk (1 liter), Loaf of Fresh White Bread (500g), Rice white (1kg), Eggs regular (12), Local Cheese (1kg), Chicken Breasts (1kg), Beef Round (1kg) (or Equivalent Back Leg Red Meat), Apples (1kg), Banana (1kg), Oranges (1kg), Tomato (1kg), Potato (1kg), Onion (1kg), Lettuce (1 head), Water (1,5 liter bottle), Bottle of Wine (Mid-Range), Domestic Beer (0,5 liter bottle), Imported Beer (0,33 liter bottle), restaurant_rating;
- **Alloggio:** Number of bed-places in tourist accommodation establishments, Total nights spent in tourist accommodation establishments, Apartment (1 bedroom) in City Centre, Apartment (1 bedroom) Outside of Centre, Apartment (3 bedrooms) in City Centre, Apartment (3 bedrooms) Outside

- of Centre, Price per Square Meter to Buy Apartment in City Centre, Price per Square Meter to Buy Apartment Outside of Centre;
- **Cultura:** Number of public libraries, Number of museum visitors, museum_rating, Cinema International Release 1 Seat;
- **Popolarità:** total tweets, Tourist Arrivals_19;
- **Tempo libero:** Cigarettes 20 Pack (Marlboro), 1 min of Prepaid Mobile Tariff Local (No Discounts or Plans), Internet (60 Mbps or More, Unlimited Data, Cable/ADSL), Fitness Club, Monthly Fee for 1 Adult, Tennis Court Rent (1 Hour on Weekend), 1 Pair of Jeans (Levis 501 Or Similar), 1 Summer Dress in a Chain Store (Zara, H&M, ecc), 1 Pair of Nike Running Shoes (Mid-Range), 1 Pair of Men Leather Business Shoes;
- **Sicurezza:** People killed in road accidents per 10000 pop, Number of murders and violent deaths.

Dato che i singoli indici inizialmente erano espressi da unità di misura differenti tra loro, si è deciso di standardizzarli per poterli confrontare con più facilità. Inoltre, è stato necessario considerare il differente verso degli indici: a questo proposito, questi ultimi sono stati modificati in modo da considerare migliori le città che presentassero i costi più bassi e i livelli di qualità più elevati per ognuno degli aspetti presi in esame.

Una volta terminate queste procedure, per creare gli indicatori generali è stata fatta la media dei singoli indici appartenenti ad ogni categoria ad eccezione della categoria “Vitto”, per cui è stata fatta prima la media di tutti gli indici relativi ai costi e, successivamente, la media tra il valore precedentemente ottenuto e l’indice `restaurant_rating`. Infine, sulla base dei valori ottenuti è stata fatta una classifica delle città per ogni aspetto ed è stato assegnato a queste ultime un punteggio da 1 a 27 in relazione alla posizione occupata: nello specifico, è stato assegnato 27 alla città che presentasse il valore più alto e 1 a quella con il valore più basso.

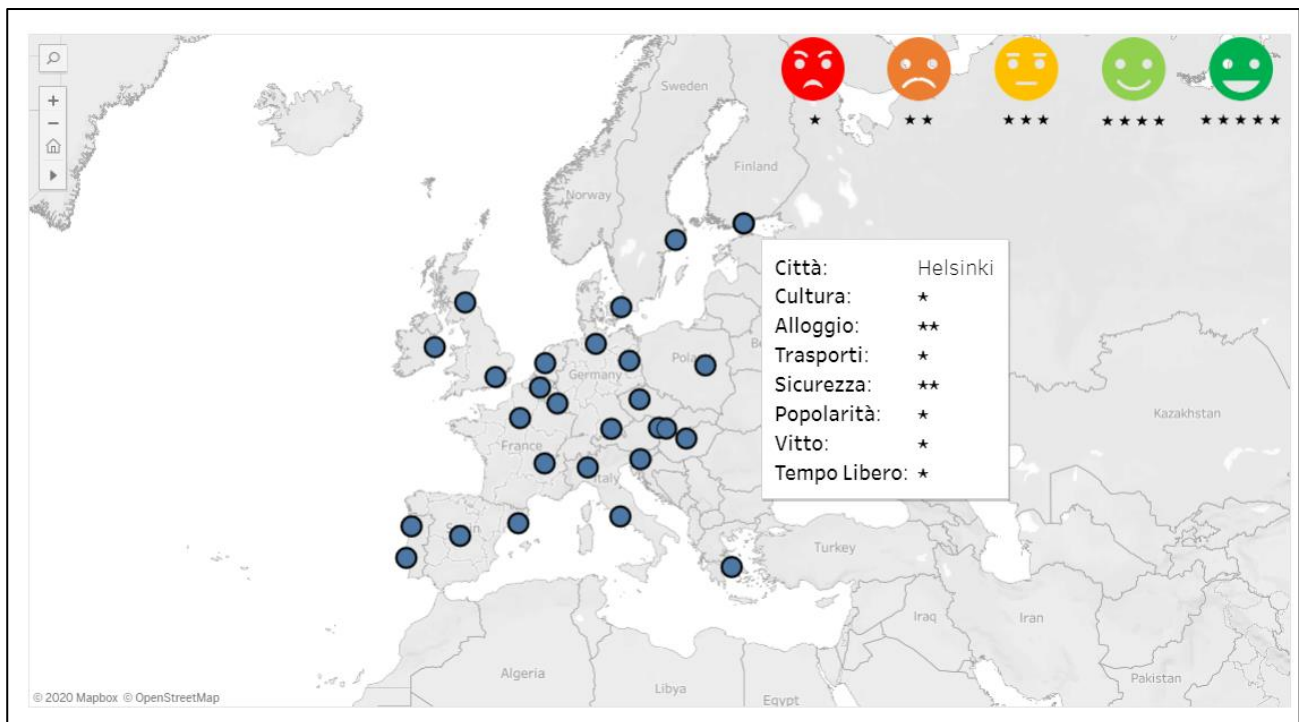
Infine, il database ottenuto è stato trasformato in un file json ed è stato caricato su Mongo.

```
{
  "_id": ObjectId("5e1f3478a30a610c90330d0b"),
  "city": "Brussels",
  "Index_Transport": 1.0641706823573387,
  "Index_Culture": 0.020911191269602798,
  "Index_Food": 0.6699972124635745,
  "Index_Housing": 0.3121238177649972,
  "Index_Popularity": -0.5925533728302532,
  "Index_Dangerousness": 0.182795900282813,
  "Index_Fun": 0.023202300941438356,
  "lat": 50.8436709,
  "lon": 4.3674366933879565
}
```

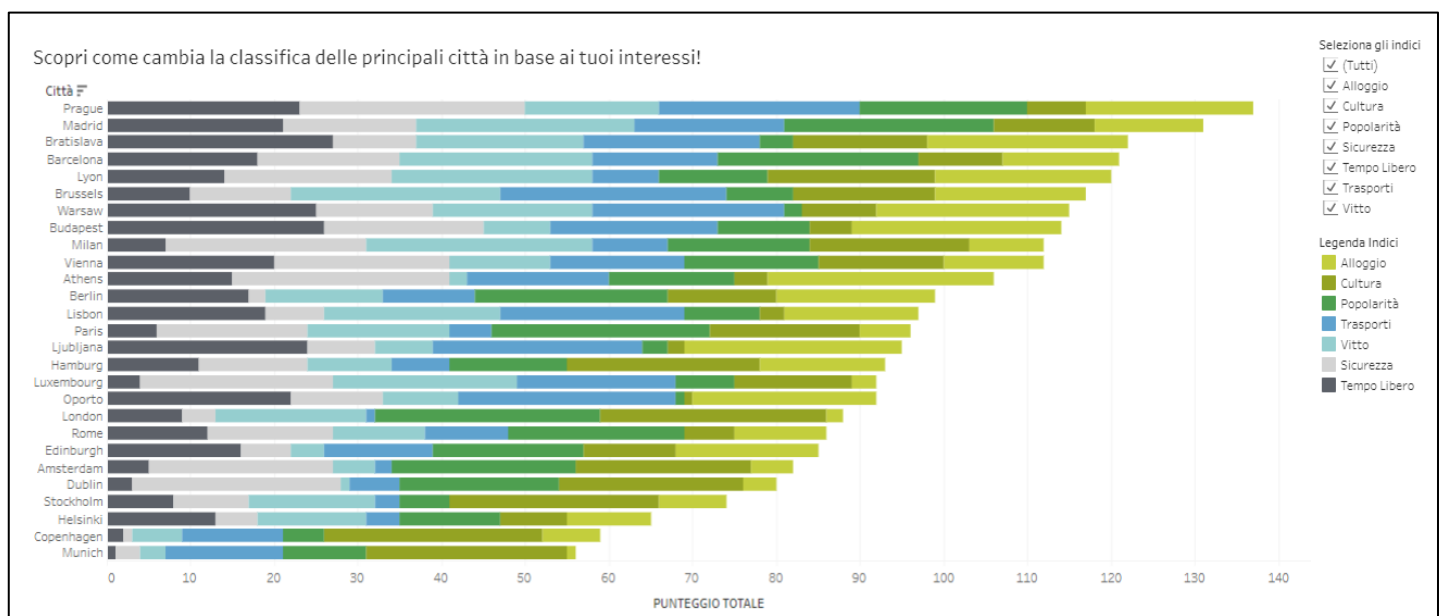
5. Visualizzazioni

Per la parte di Data Visualization è stato utilizzato il software Tableau. Il database utilizzato è stato quello ottenuto nel capitolo precedente, contenente le 27 città scelte, i relativi dati di latitudine e longitudine e i valori di tutti gli indicatori generali calcolati in precedenza.

La **prima visualizzazione** permette di avere una visione generale delle principali città europee sotto ognuno dei 7 aspetti considerati. Nella mappa le città vengono contrassegnate da un pallino blu. Cliccando su ognuno di essi viene visualizzato il punteggio ottenuto per ogni indicatore utilizzando delle piccole stelle nere: come mostrato anche nella legenda, ad 1 stella corrisponde il minimo mentre a 5 stelle corrisponde il massimo.



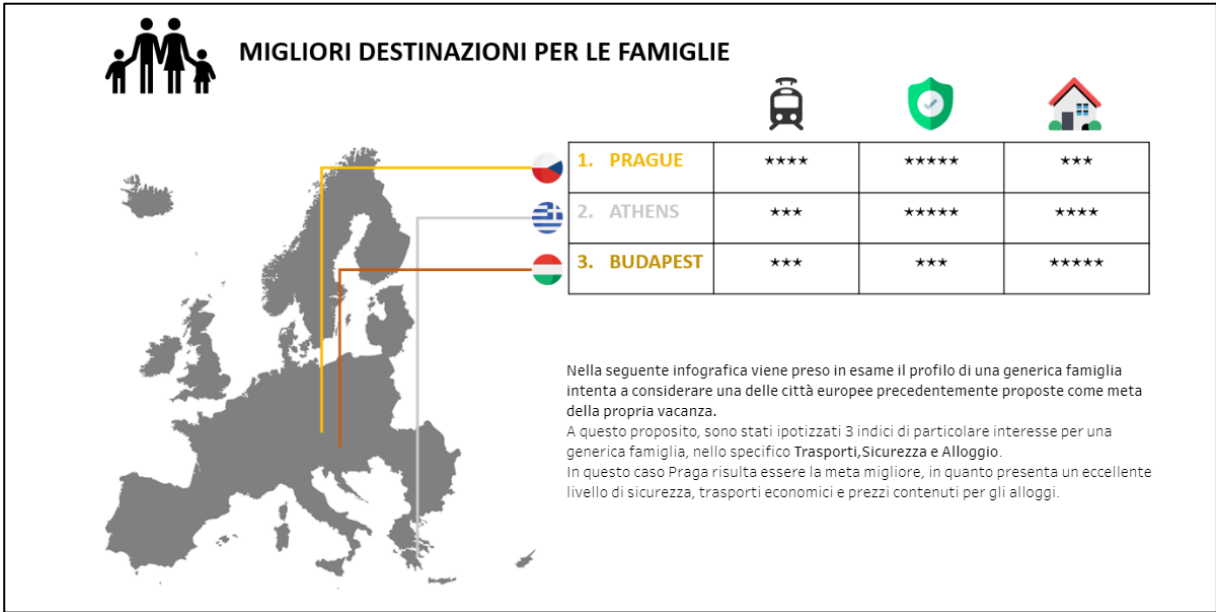
La **seconda visualizzazione** permette di vedere la classifica delle principali città europee sulla base degli indicatori scelti. Per ogni indicatore viene preso in considerazione il sistema di punteggi precedentemente descritto (da 1 a 27). La classifica visualizzata varia in funzione della somma dei punteggi assegnati alle città, sulla base dell'indicatore o degli indicatori selezionato/i.



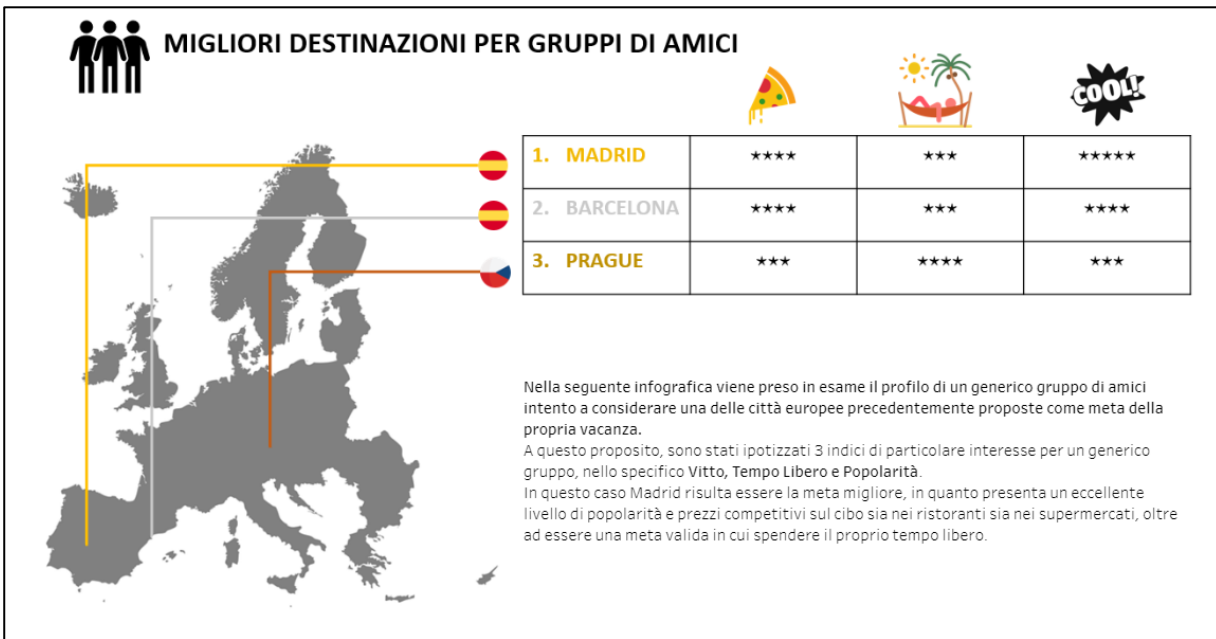
Nella **terza e nella quarta visualizzazione** sono state proposte due possibili tipologie di turisti interessati a valutare le città sulla base dello studio presentato.

Nel primo caso, viene presa in esame un'ipotetica famiglia a cui vengono suggeriti Trasporti, Sicurezza e Alloggio come indicatori generali per valutare meglio la città più adatta in cui andare.

In questo caso Praga risulta essere la meta migliore, in quanto presenta un eccellente livello di sicurezza, trasporti economici e funzionali e una buona offerta per quanto riguarda gli alloggi.



Nel secondo caso, invece, viene presa in esame un ipotetico gruppo di amici a cui vengono suggeriti Vitto, Tempo libero e Popolarità come indicatori generali per valutare meglio la città più adatta in cui andare. In questo caso Madrid risulta essere la meta migliore, in quanto presenta un eccellente livello di popolarità e prezzi competitivi sul cibo sia nei ristoranti sia nei supermercati, oltre ad essere una meta valida in cui spendere il proprio tempo libero.



Per avere una miglior visualizzazione, le infografiche sono disponibili al seguente [link](#).

6. Quality Assessment

6.1 Test Euristico

Di seguito si riportano i punti critici osservati dalle 7 persone coinvolte nel test:

1. La barra di ricerca delle città nella prima infografica risulta essere di dimensione troppo ridotta e non viene notata facilmente dall'utente;
2. La selezione degli indicatori nella seconda infografica risulta poco visibile;
3. L'icona inizialmente utilizzata per l'indicatore di popolarità non ne fa intuire il significato con facilità.

Inoltre, si è notato che gli utenti iniziavano ad interagire con l'infografica prima di aver letto le rispettive istruzioni, cercando di individuare le città di interesse.

6.2 User Test

Le domande poste agli intervistati sono state quattro, due relative alla prima infografica e due relative alla seconda. Di seguito si riportano il testo di ogni domanda e la relativa risposta:

- **INFOGRAFICA 1:**

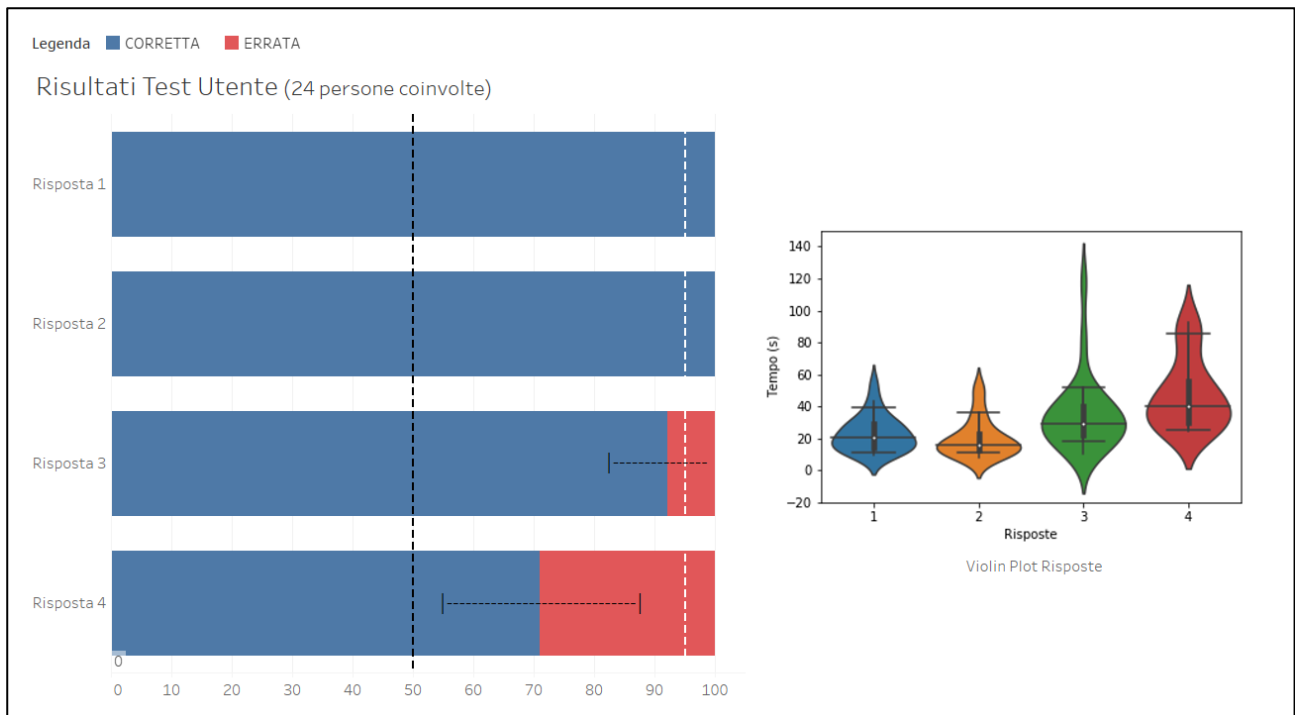
1. Qual è il numero di stelle della città "Dublin" sull'indicatore "Vitto"?
Risposta: ★
2. Su quale indicatore ha il maggior numero di stelle la città "Stockholm"?
Risposta: Cultura

- **INFOGRAFICA 2:**

1. Qual è la città migliore secondo l'indicatore "Sicurezza"?
Risposta: Prague (Praga)
2. Qual è la città peggiore secondo la combinazione degli indicatori "Cultura", "Tempo libero" e "Trasporti"?
Risposta: Helsinki

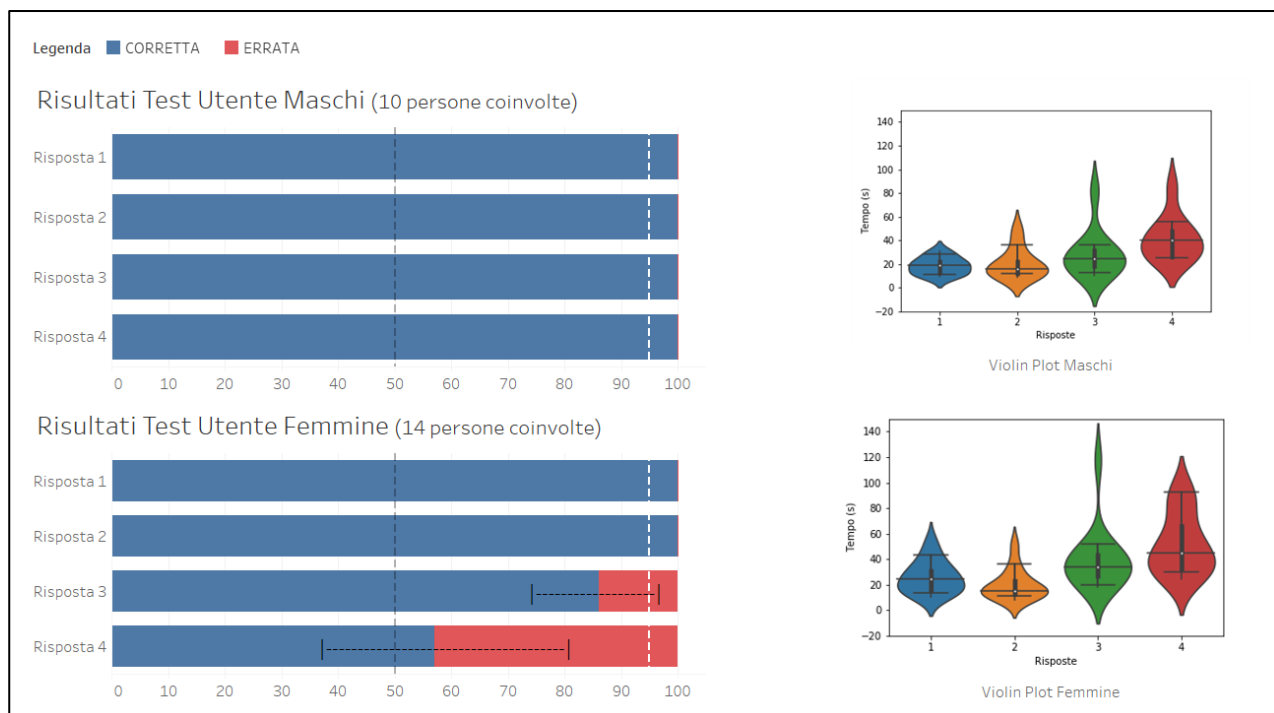
Le risposte relative alle prime due domande erano reperibili cliccando sul pallino corrispondente alla città e valutando il numero di stelle per ogni indicatore. Le risposte alle seconde due domande erano reperibili selezionando gli indicatori richiesti e osservando la classifica generata nella visualizzazione.

Di seguito si riportano i risultati del test utente sia per il campione totale, sia per il campione diviso tra maschi e femmine.



Le domande sono state scelte appositamente di bassa difficoltà per verificare se le infografiche fossero immediate ed intuitive. Gli utenti intervistati sono stati 24, tutti studenti, di cui 14 di sesso femminile e 10 di sesso maschile. Si è riscontrato che alle prime due domande tutti gli utenti hanno risposto correttamente, impiegando in media circa 20 secondi (nello specifico, il 90% degli utenti tra i 15 e i 25 secondi). Nella terza domanda il tasso di errore è stato contenuto, intorno all'8% ma, dato l'intervallo di confidenza del tasso di errore, non si può affermare che quest'ultimo sia significativamente superiore al 5%. In questa risposta si è registrato il più alto tempo singolo di risposta e il 90% degli intervistati ha impiegato tra i 25 e i 45 secondi. Nell'ultima domanda si è registrato il tasso di errore più alto, circa 29%: tenendo conto dell'intervallo di confidenza, questo tasso risulta essere significativamente maggiore del 5%. Il tempo di risposta medio è stato il più elevato, circa 50 secondi, ed il 90% degli intervistati ha impiegato tra i 40 ed i 60 secondi.

In secondo luogo, sono stati analizzati i risultati dividendo gli utenti per genere per verificare eventuali differenze (come mostrato nella figura sottostante).



Dai risultati si può notare come gli intervistati di sesso maschile abbiano risposto correttamente a tutte le domande, mentre gli intervistati di sesso femminile abbiano commesso qualche errore, soprattutto nella quarta domanda, nella quale non si può stabilire se siano significativamente maggiori le risposte esatte rispetto a quelle sbagliate.

Visti i numerosi errori commessi nella quarta domanda, sono state controllate tutte le risposte relative al suddetto quesito: circa il 15% degli intervistati ha risposto indicando ‘Bratislava’ come la città avente ottenuto il punteggio peggiore secondo la combinazione degli indicatori “Cultura”, “Tempo libero” e “Trasporti”, mentre è quella che ha il punteggio maggiore sulla somma di questi 3 indicatori. Quindi, si può pensare che i numerosi errori commessi siano dovuti ad una lettura frettolosa della domanda piuttosto che ad un’incomprensione dell’infografica.

In conclusione, possiamo affermare che le visualizzazioni sono state comprese dalla maggior parte degli intervistati.

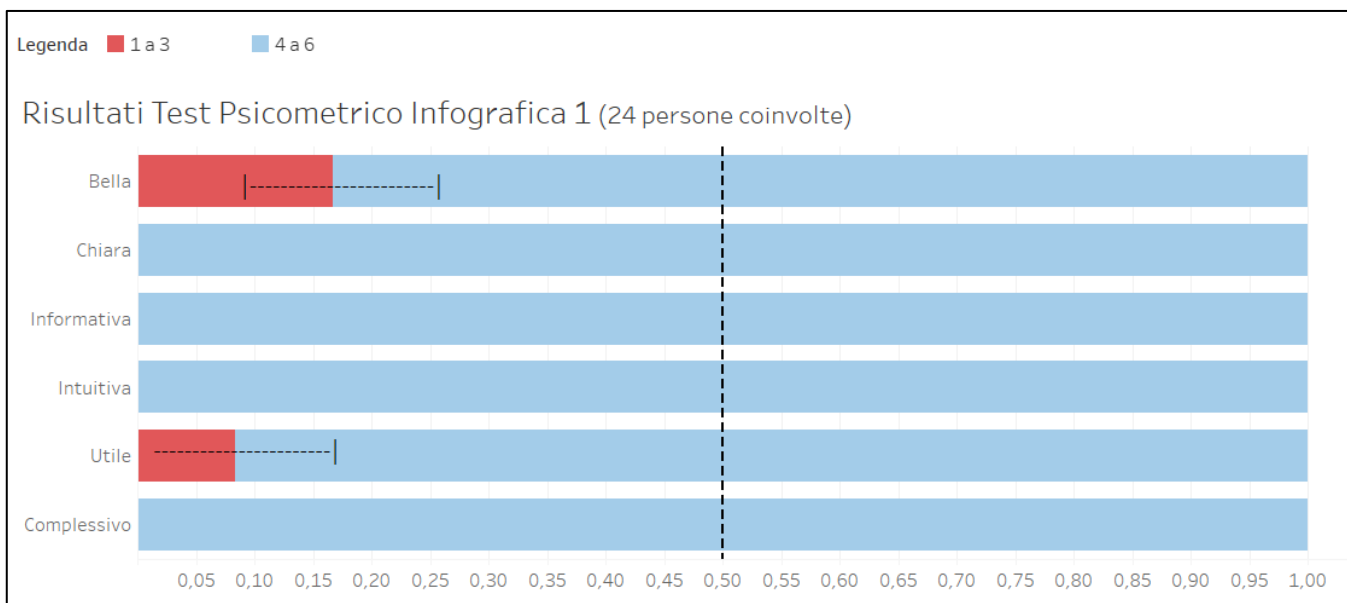
6.1 Test Psicometrico

Al termine dell'intervista è stato chiesto ai 24 intervistati di rispondere al questionario seguente:

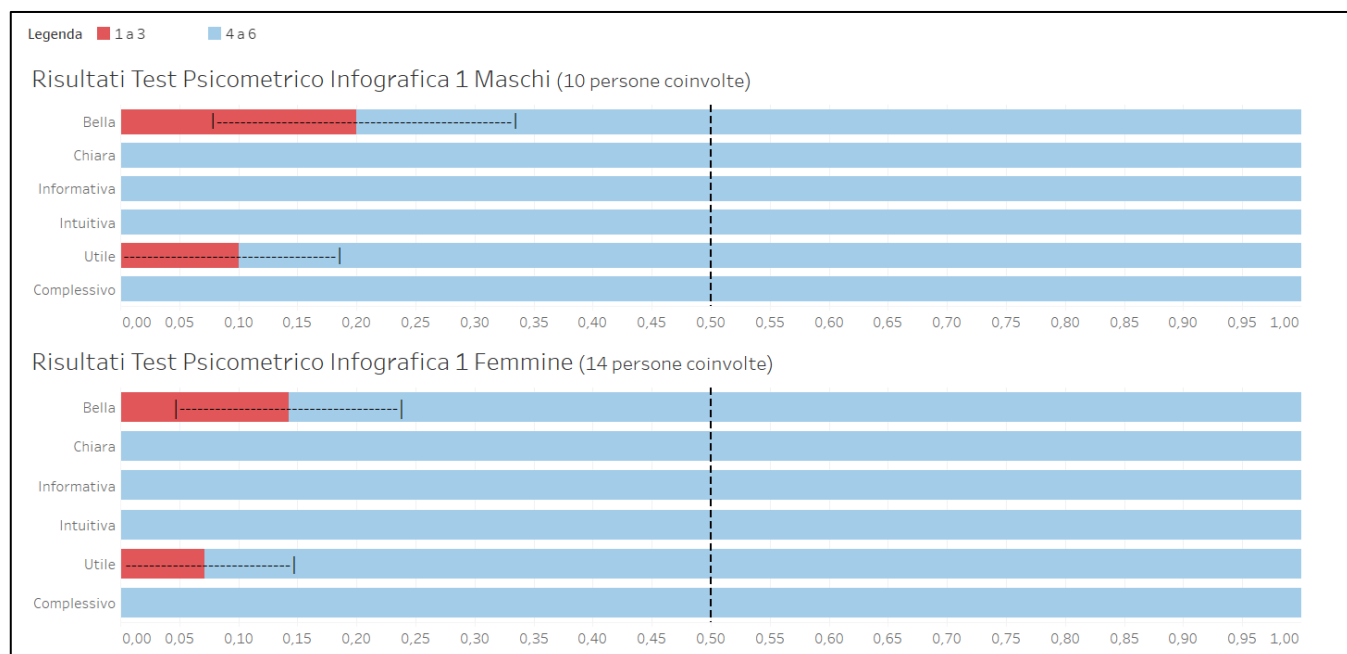
Valuta la qualità dell'infografica dando un valore da 1 (pochissimo) a 6 (moltissimo) a ciascuno dei seguenti aspetti: *						
	1 - pochissimo	2	3	4	5	6 - moltissimo
Utile	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Chiara	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Informativa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Bella	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Intuitiva	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Valuta l'infografica indicando un valore di qualità complessivo da te percepito: *						
	1 - bassissimo	2	3	4	5	6 - altissimo
Valore complessivo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

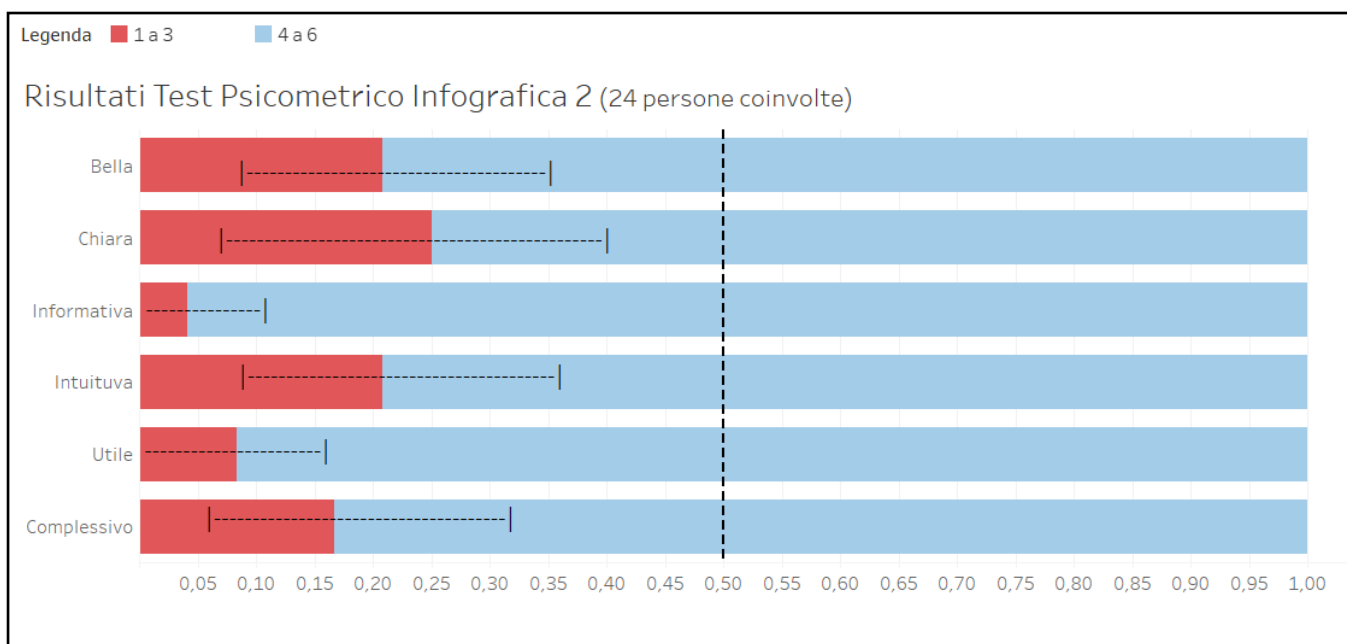
Le risposte sono state divise in due gruppi: da 1 a 3 e da 4 a 6, distinguendo in questo modo le valutazioni negative da quelle positive. Di seguito vengono riportati i risultati ottenuti per le due infografiche, mostrando prima i dati totali, poi i dati suddivisi per genere.



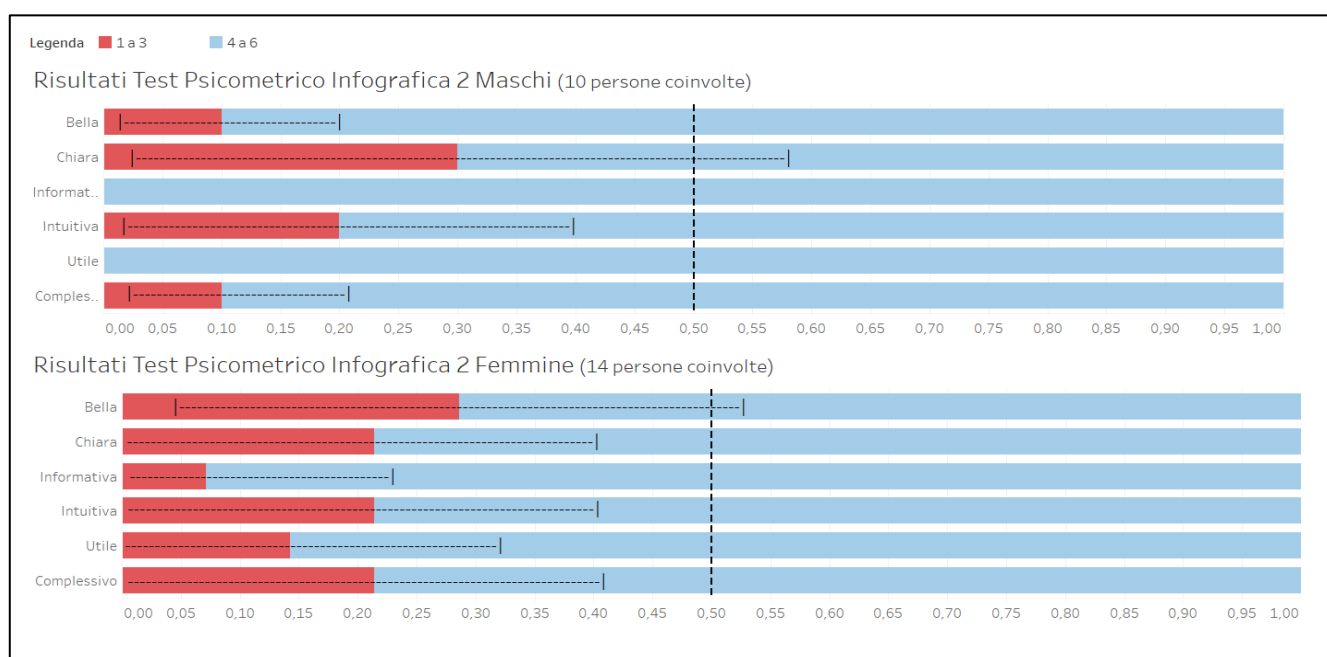
Dai risultati del questionario relativo alla prima infografica si può affermare che le valutazioni positive siano state significativamente maggiori rispetto a quella negative in tutte le categorie. Nella figura in basso, invece, vengono mostrati i risultati relativi alla prima infografica, divisi tra maschi e femmine.



Nonostante la divisione in generi, non si notano particolari differenze di andamento rispetto al grafico sulle valutazioni totali.

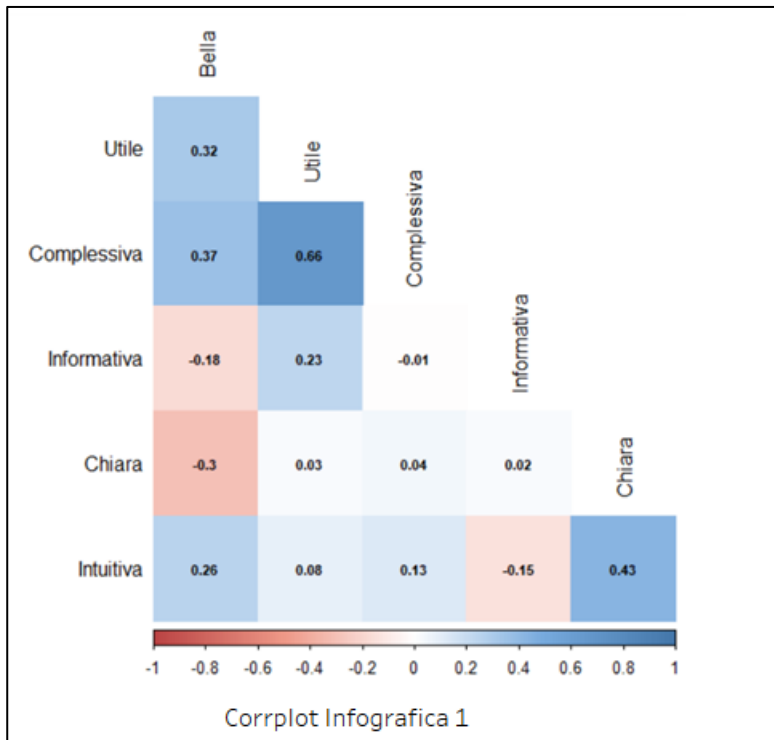


Anche per quanto riguarda le valutazioni relative alla seconda infografica si può affermare che quelle positive siano state significativamente maggiori rispetto a quella negative in tutte le categorie considerate.

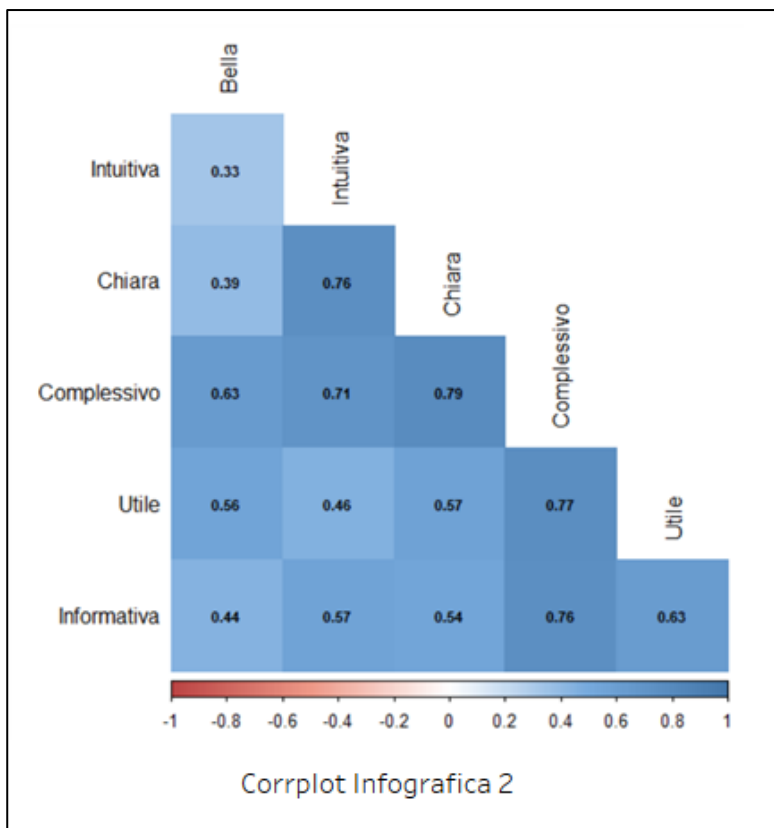


In questo caso, nonostante una differenza più netta tra le valutazioni di maschi e femmine, rimangono preponderanti i voti positivi rispetto a quelli negativi.

Infine, vengono mostrati due grafici che rappresentano la correlazione tra le categorie di scelta utilizzate nel test.



Dal grafico che mostra le correlazioni per la prima infografica risulta che l'unica correlazione medio/alta tra le risposte è tra 'Utile' e 'Complessiva'.



Dal grafico che mostra le correlazioni per la seconda infografica risulta che la variabile ‘Complessiva’ presenta una correlazione medio/alta con tutte le altre categorie. Inoltre, un’altra correlazione dal valore elevato risulta essere quella tra le modalità ‘Chiara’ e ‘Intuitiva’.

Per avere una miglior visualizzazione, i risultati dei test sono disponibili al seguente [link](#).

7. Conclusioni e Sviluppi futuri

Il quesito di partenza dello studio è stato uno solo: “*Vacanza in città: dove andare?*”. Tuttavia, a questo quesito non si può rispondere univocamente. Lo scopo, infatti, era quello di supportare un potenziale turista nella scelta di una meta per una futura vacanza e proprio per questo è stata data la possibilità al turista di selezionare le categorie più indicate per valutare le città.

Le prime due infografiche forniscono informazioni utili per l’utente e lo rendono in grado di confrontare le città sia sulla base di ogni singolo indicatore, sia sulla base di diverse combinazioni tra gli indicatori stessi. Infine, le ultime due visualizzazioni forniscono due ipotetiche risposte alla domanda iniziale, tenendo in considerazione due diversi profili di turisti, con una diversa scelta di indicatori.

Tuttavia, nonostante un’ampia scelta di variabili iniziali, lo studio presenta alcuni punti deboli, primo fra tutti il fatto che molte fra le città prese in esame inizialmente presentassero valori mancanti su alcuni indici (fatto che ha portato al confronto di sole 27 città). Inoltre, alcuni valori non sono reali, ma ricavati tramite previsioni per l’anno considerato. Infine, la raccolta dei tweet georeferenziati in cui vengono citate le città non è sempre ricollegabile al concetto di popolarità turistica: la popolarità in questione, infatti, può essere considerata tenendo conto di diversi ambiti, dalla politica allo sport, dalle notizie di gossip a quelle di cronaca, motivo per il quale non si può definire completamente affidabile. Tutti questi problemi rendono l’analisi incompleta e non sempre realistica per quanto concerne la descrizione delle diverse città.

Ai fini dello studio, invece, il fatto di non avere un’univoca risposta non rappresenta un punto debole, bensì un punto di inizio per migliorare il lavoro svolto. Infatti, disporre di un sempre maggior numero di dati, ottenuti in tempo reale, relativi ad un sempre maggior numero di città permetterebbe di avere un quadro sempre più chiaro della situazione qualitativa di numerose mete turistiche europee e mondiali e fornirebbe un aiuto sempre più concreto agli utenti per rispondere in modo univoco alla domanda iniziale. Infatti, in un mondo sempre più globalizzato come quello in cui viviamo oggi, dove viaggiare non è mai stato così semplice, risulta molto utile scegliere la meta per le proprie vacanze basandosi su differenti punti di vista.