

GUIDA ALL'UTILIZZO DEL CODICE E ALLA REPLICAZIONE DEL PROGETTO

1. ACQUISIZIONE DATI

Nella prima parte del progetto abbiamo reperito i dati dai siti di Eurostat, Kaggle ed Euromonitor e abbiamo salvato i seguenti file: 'TA_restaurants_curated.csv', 'Hotel_Reviews.csv', 'cultura_e_turismo.csv', 'popolazione.csv', 'ambiente.csv', 'condizioni_vita.csv', 'trasporti.csv', 'tripadvisor_museum.json', 'numero_turisti.csv'.

Tramite l'utilizzo dei seguenti script, 'Tripadv.ipynb' e 'Eurostat.ipynb', abbiamo selezionato le variabili e le osservazioni di nostro interesse e abbiamo caricato i dataset visibili su Mongo.

Attraverso lo script 'ScrapingNumbeo.ipynb', sono stati ottenuti i dati dal sito di Numbeo tramite scraping.

Per quanto riguarda i dati di Twitter, abbiamo utilizzato gli script 'ProducerTweet.ipynb' e 'ConsumerTweet.ipynb' per scaricarli, simulare la velocità tramite KAFKA e inserire il dataset su Mongo. Lo script 'ProducerTweet.ipynb' deve essere aggiornato manualmente per poter reperire i Tweet riguardanti tutti i mesi dell'anno. Tramite lo script 'TwitterIntegration.ipynb' è stata creata la nuova colonna di interesse e il dataset finale è stato successivamente caricato su Mongo.

2. DATA INTEGRATION

Nel seguente script, 'DataIntegration.ipynb', è stato riportato il codice utilizzato per l'integrazione di tutti i diversi dataset ottenuti e la creazione del dataset finale.

3. CREAZIONE DEGLI INDICATORI

Tramite lo script, 'Indicators.ipynb', sono stati creati i sette indicatori finali che sono stati utilizzati per realizzare le visualizzazioni.