

Progetto analisi dataset con Python

Selezione del Dataset

Il dataset deve essere in formato JSON, provenire da un portale di open data e ricco di informazioni per supportare analisi.

Se il dataset è accessibile tramite un link diretto stabile (JSON/GeoJSON), è possibile utilizzare direttamente l'URL pubblico.

In alternativa, il dataset può essere scaricato e caricato su un repository GitHub, mantenendo il formato originale e citando la fonte.

In entrambi i casi, il dataset deve essere caricato in Google Colab tramite richiesta HTTP (`requests.get()`), utilizzando un link diretto al file JSON (link Raw nel caso di GitHub).

Analisi dei Dati

Eseguire l'analisi del dataset utilizzando le tecniche viste in laboratorio e, se opportuno, ulteriori analisi aggiuntive.

L'analisi può includere, a titolo di esempio:

- esplorazione della struttura del dataset (numero di record, campi, tipi di dato);
- selezione e filtraggio dei dati;
- conteggi, frequenze e aggregazioni;
- calcolo di valori statistici di base (min, max, media, ecc.);
- confronto tra sottoinsiemi di dati.

Per ogni operazione svolta, fornire una descrizione chiara del procedimento e un commento sui risultati ottenuti.

In alternativa o in aggiunta, è possibile utilizzare una libreria di visualizzazione interattiva (es. Plotly, Bokeh, Altair), motivando la scelta e spiegando cosa l'interattività aggiunge alla lettura dei dati.

Visualizzazioni Grafiche*

<https://matplotlib.org/stable/gallery/index>

Utilizzando la libreria matplotlib, realizzare almeno **tre grafici**, ciascuno scelto per rappresentare un tipo diverso di informazione, ad esempio:

- andamento nel tempo,
- confronto tra categorie,
- relazione tra due variabili o distribuzione dei dati.

Per ogni grafico indicare che tipo di informazione rappresenta.

Creazione di Mappe Interattive

Nel caso in cui il dataset contenga informazioni geografiche (coordinate o località), è auspicabile l'uso della libreria Folium per realizzare una visualizzazione su mappa.

<https://python-visualization.github.io/folium/latest/>

Aggiungere una spiegazione dettagliata per ogni mappa e per gli elementi visualizzati.

Esposizione Finale

Al termine del progetto, condividere il link di Google Colab.

Durante l'esposizione finale, gli studenti illustreranno il ragionamento seguito in ciascuna fase dell'analisi, discuteranno le scelte fatte e risponderanno alle domande del docente.

CONSEGNA:

1. Inserire nella tabella il link del dataset scelto.
 2. Al termine del progetto consegnare il link al progetto svolto su Google Colab.
In alternativa, consegnare il repository Github per chi usa VSC o Pycharm.
-

* Regola pratica per il tipo di grafico

Come cambia nel tempo? line plot (o area)

Chi è più grande/piccolo tra categorie? bar chart

C'è relazione tra due variabili? scatter plot (con trendline se serve)

Com'è distribuito un valore? histogram (o boxplot/violin)

Composizione sul totale? spesso stacked bar.