

Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente

**LINK COLAB:** https://colab.research.google.com/#scrollTo=A9-KYPDpSnlB

**SVOLGIMENTO**

# Analisi diffusione COVID-19 nel mondo Il committente richiede di avere un report su casi e vaccinazioni in diverse aree del mondo;

# a tal fine, richiede di utilizzare il dataset, curato da Our World in Data, all'indirizzo https://github.com/owid/covid-19-data/tree/master/public/data; alla voce "🗂 dataset" scaricare il dataset nel formato che si preferisce.

import pandas as pd

# Si richiede di verificare le dimensioni del dataset e i relativi metadati;

df=pd.read\_csv("owid-covid-data.csv")

df.shape

df.info()

# Si chiede poi per ogni continente di trovare il numero di casi totali avvenuti in quello stesso continente; si chiede di non considerare eventuali locazioni che nel dataset non appartengono ad alcun continente;

df\_continent\_total\_cases=df.groupby("continent")["total\_cases"].sum()

print(df\_continent\_total\_cases)

# Sempre riguardo i casi totali, si chiede di scrivere del codice che,date due variabili contenenti i nomi di continenti,ne confronti i seguenti relativi descrittori statistici:valore massimo, media, e percentuale rispetto al numero dei casi totali nel mondo (calcolati anche sulle locazioni senza indicazione di continente);

def compare\_continents(continents):

  for continent in continents:

        continent\_data = df[df['continent'] == continent]['total\_cases']

        # Calcolo i descrittori statistici per il continente

        continent\_max = continent\_data.max()

        continent\_mean = continent\_data.mean()

        continent\_perc = (continent\_data.sum() / df['total\_cases'].sum()) \* 100

        # Stampo i risultati

        print(f"Statistiche per {continent}:")

        print(f"Massimo: {continent\_max}")

        print(f"Media: {continent\_mean}")

        print(f"Percentuale: {continent\_perc:.2f}%")

        print()

# Utilizzo la funzione per confrontare i continenti

compare\_continents(["Africa", "Asia"])

# Selezionare i dati relativi all'Italia nel 2022, e mostrare con un grafico adeguato l'evoluzione del casi totali rispetto alla data. Mostrare poi con un grafico adeguato il numero di nuovi casi rispetto alla data (filtrare i dati se necessario). Mostrare infine un grafico che mostra l'andamento della somma cumulativa nuovi casi del 2022, e commentare se ci sono similarità con l'andamento del numero dei casi totali;

italy\_2022 = df[(df['location'] == 'Italy') & (df['date'].str.startswith('2022'))]

# Plot total cases in base alla data

italy\_2022.plot(x='date', y='total\_cases', title='Total Cases in Italy (2022)')

# Plot new cases per il periodo

italy\_2022.plot(x='date', y='new\_cases', title='New Cases in Italy (2022)')

# Plot cumulative domma nuovi casi

italy\_2022['cumulative\_new\_cases'] = italy\_2022['new\_cases'].cumsum()

italy\_2022.plot(x='date', y='cumulative\_new\_cases', title='Cumulative New Cases in Italy (2022)')

Ci sono molte similarità tra il grafico dell'andamento del numero dei casi totali in relazione alla somma cumulativa nuovi casi del 2022

# Riguardo le nazioni di Italia, Germania e Francia, mostrare in un boxplot la differenza tra queste nazioni riguardo il numero di pazienti in terapia intensiva (Intensive Care Unit, ICU) da maggio 2022 (incluso) ad aprile 2023 (incluso), e scrivere un breve commento a riguardo;

icu\_data = df[(df['location'].isin(['Italy', 'Germany', 'France'])) &

                      (df['date'] >= '2022-05-01') & (df['date'] <= '2023-04-30')]

icu\_data.boxplot(column='icu\_patients', by='location', grid=False)

# Riguardo le nazioni di Italia, Germania, Francia e Spagna nel 2023, mostrare la somma dei pazienti ospitalizzati per ognuna -- se ci sono dati nulli, suggerire se può essere possibile gestirli tramite sotituzione o meno

hospital\_data\_2023 = df[(df['location'].isin(['Italy', 'Germany', 'France', 'Spain'])) &

                                (df['date'].str.startswith('2023'))]

hospitalized\_sum = hospital\_data\_2023.groupby('location')['hosp\_patients'].sum()

print(hospitalized\_sum)

In questo arco di tempo, in Germania non sarà stato registrato nessun paziente ospedalizzato,o bisogna leggere il motivo di tutti questi NaN