Name:Premkumar Garud

Roll No.:CS5-90

PRN:202401100001

Dataset:Covid-19 Dataset

import pandas as pd

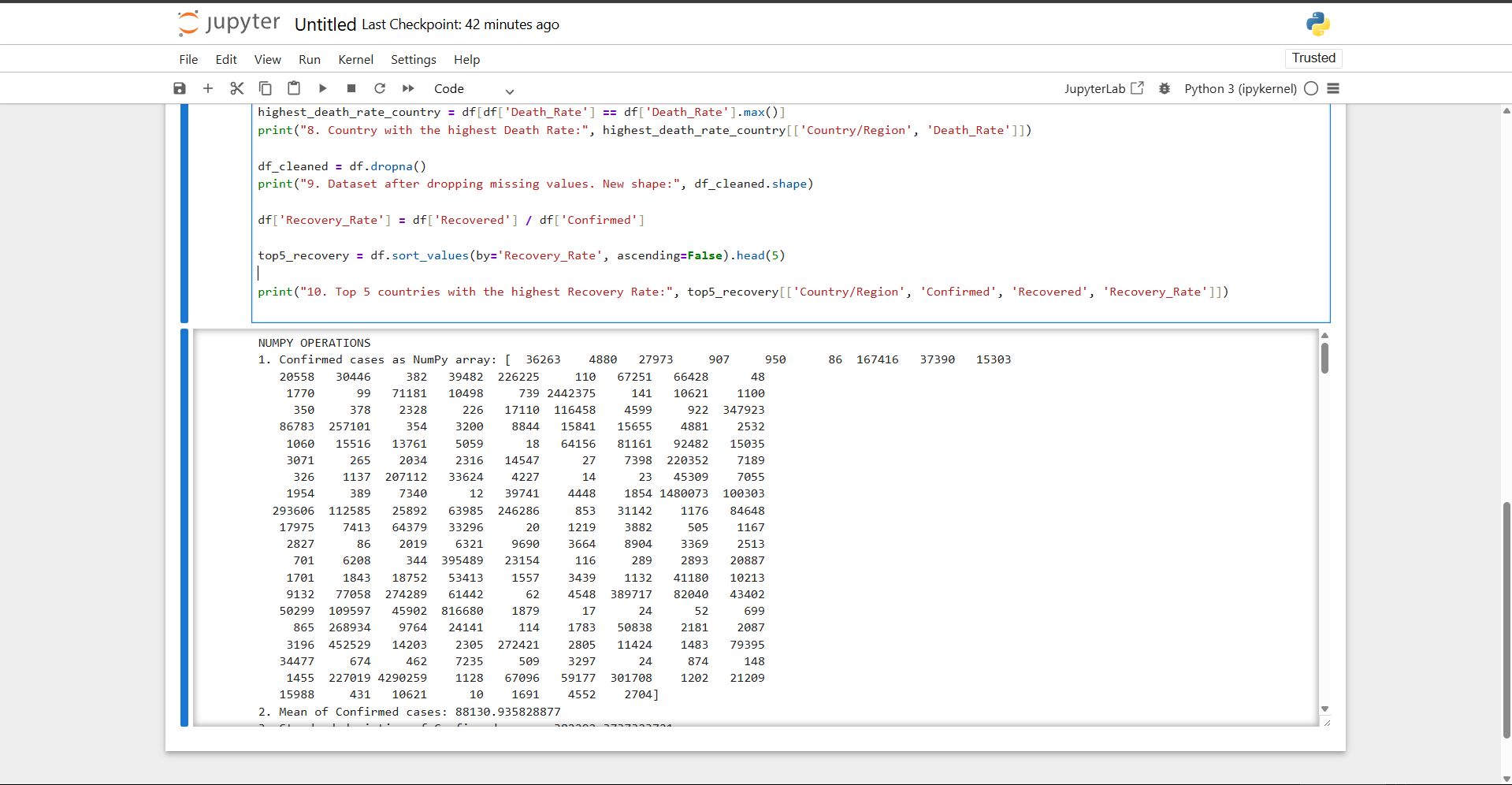
import numpy as np

df = pd.read\_csv('country\_wise\_latest.csv')

print("NUMPY OPERATIONS")

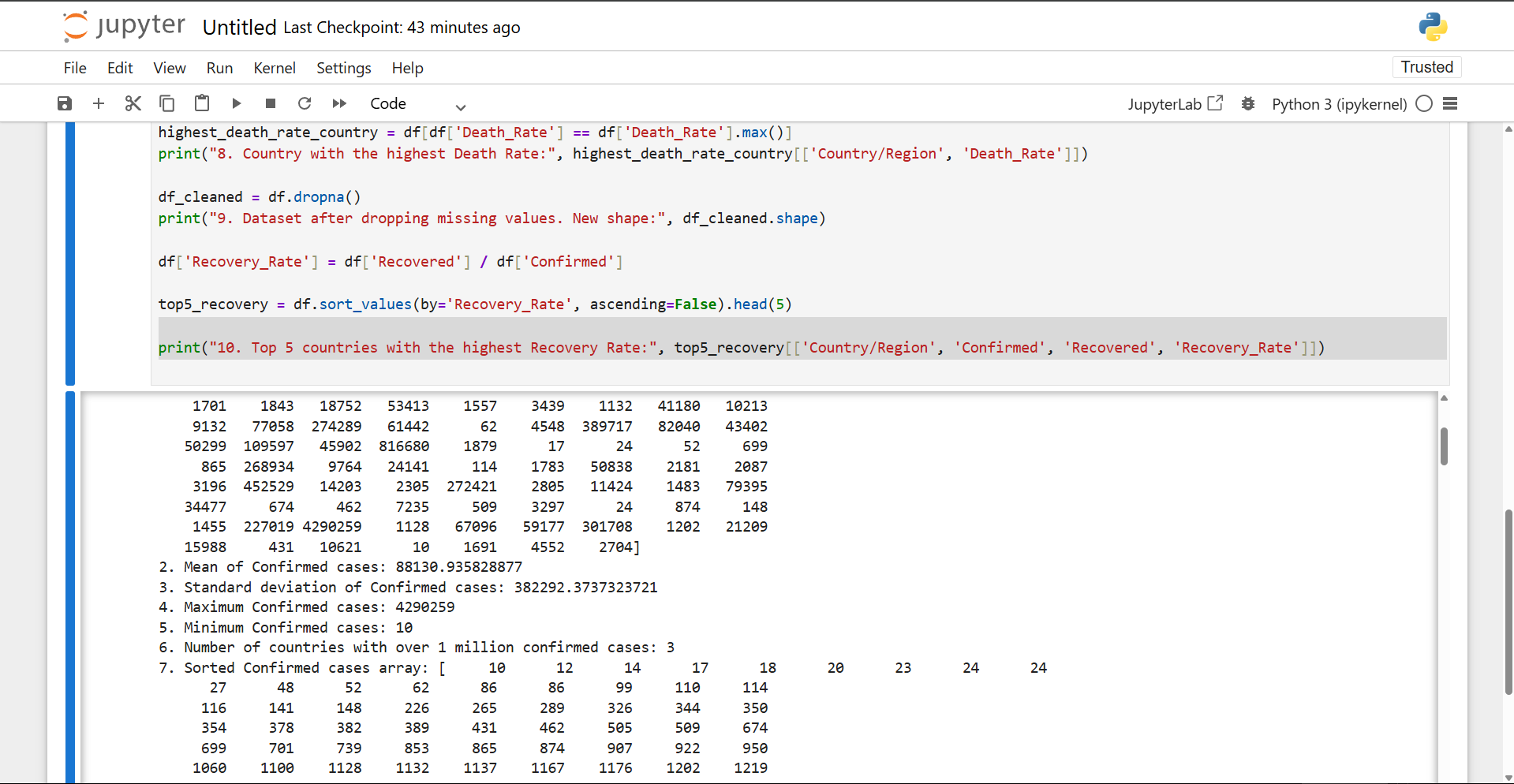
confirmed\_array = df['Confirmed'].to\_numpy()

print("1. Confirmed cases as NumPy array:", confirmed\_array)

Output: 

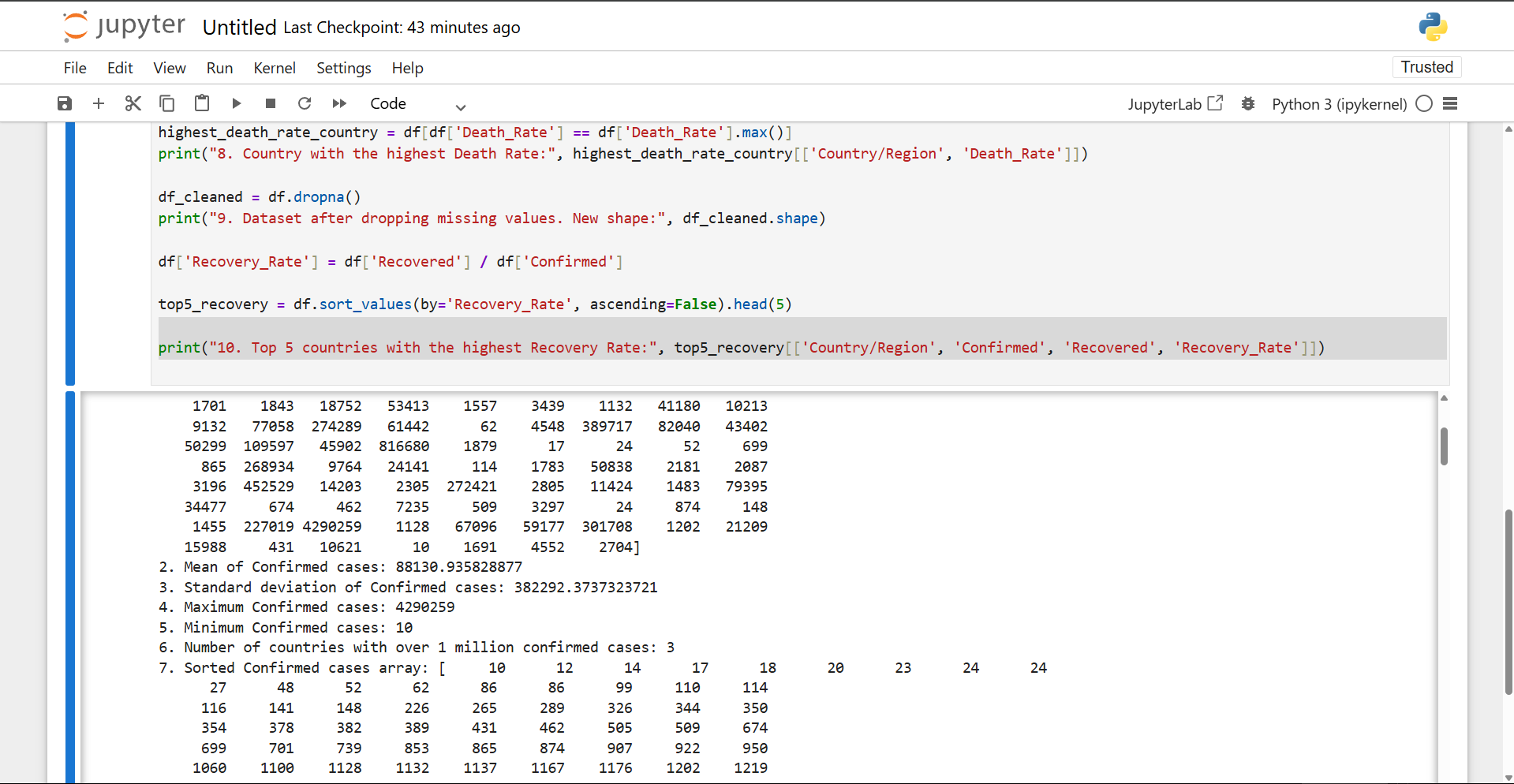
mean\_confirmed = np.mean(confirmed\_array)

print("2. Mean of Confirmed cases:", mean\_confirmed)

Output: 

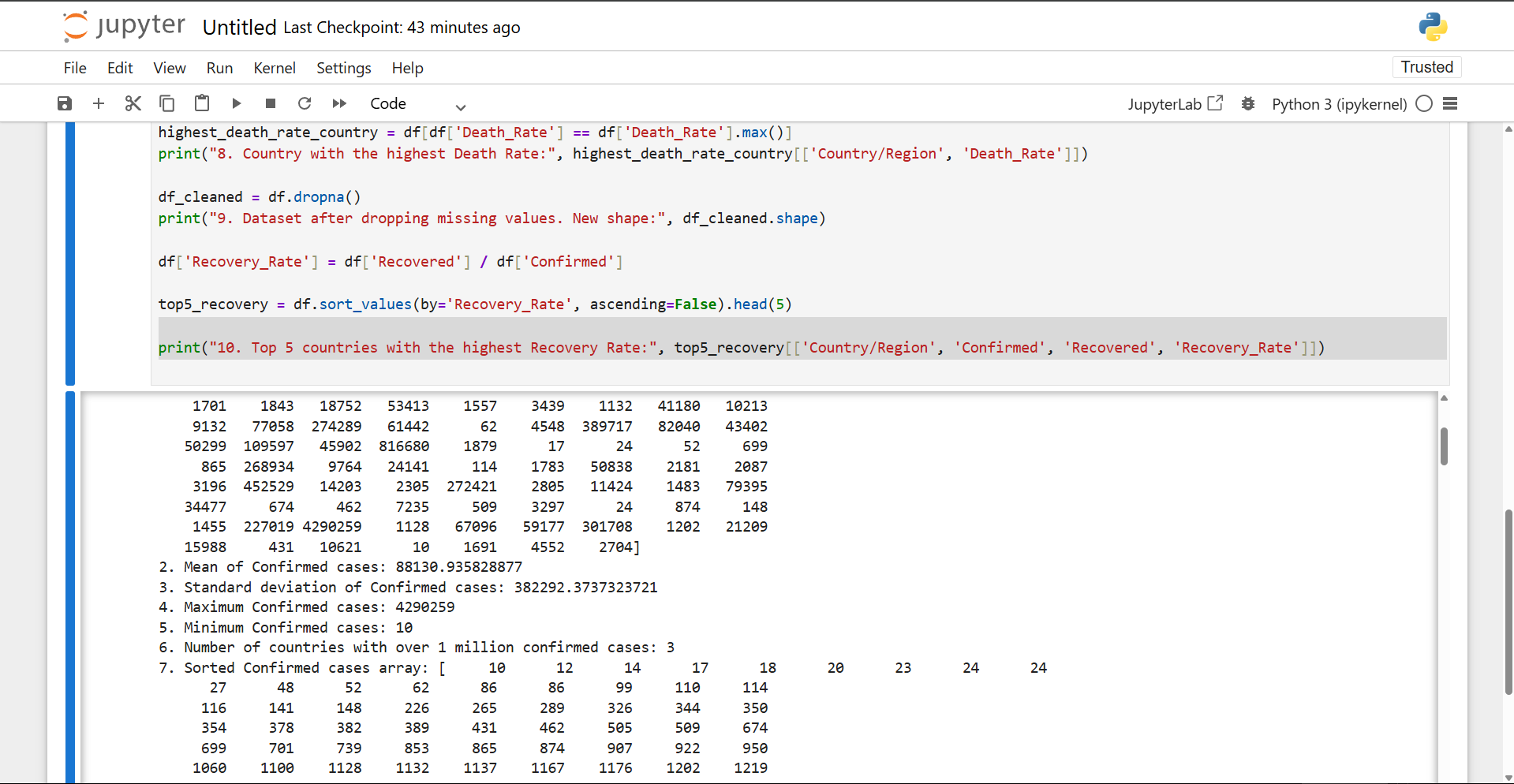
std\_confirmed = np.std(confirmed\_array)

print("3. Standard deviation of Confirmed cases:", std\_confirmed)

Output: 

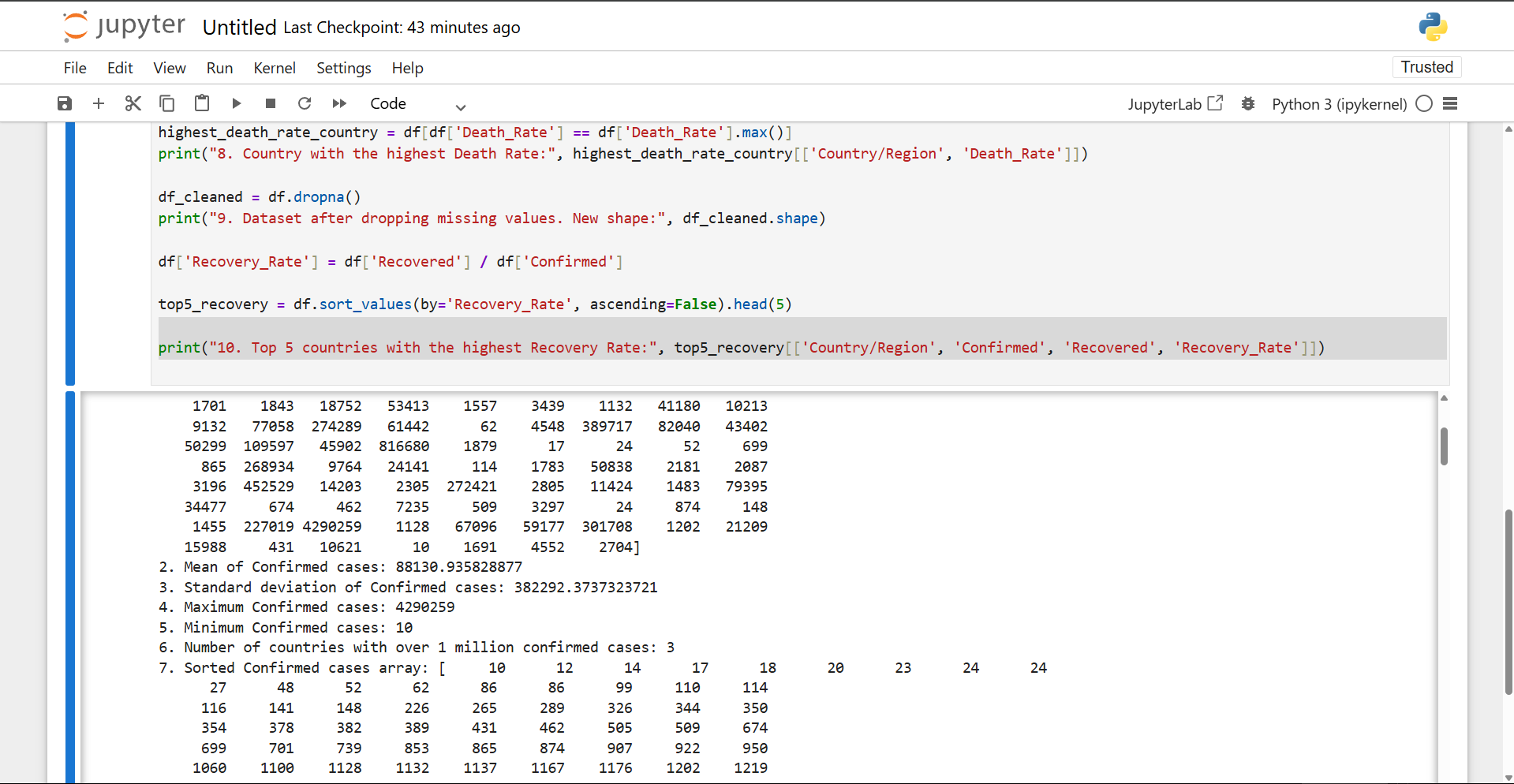
max\_confirmed = np.max(confirmed\_array)

print("4. Maximum Confirmed cases:", max\_confirmed)

Output: 

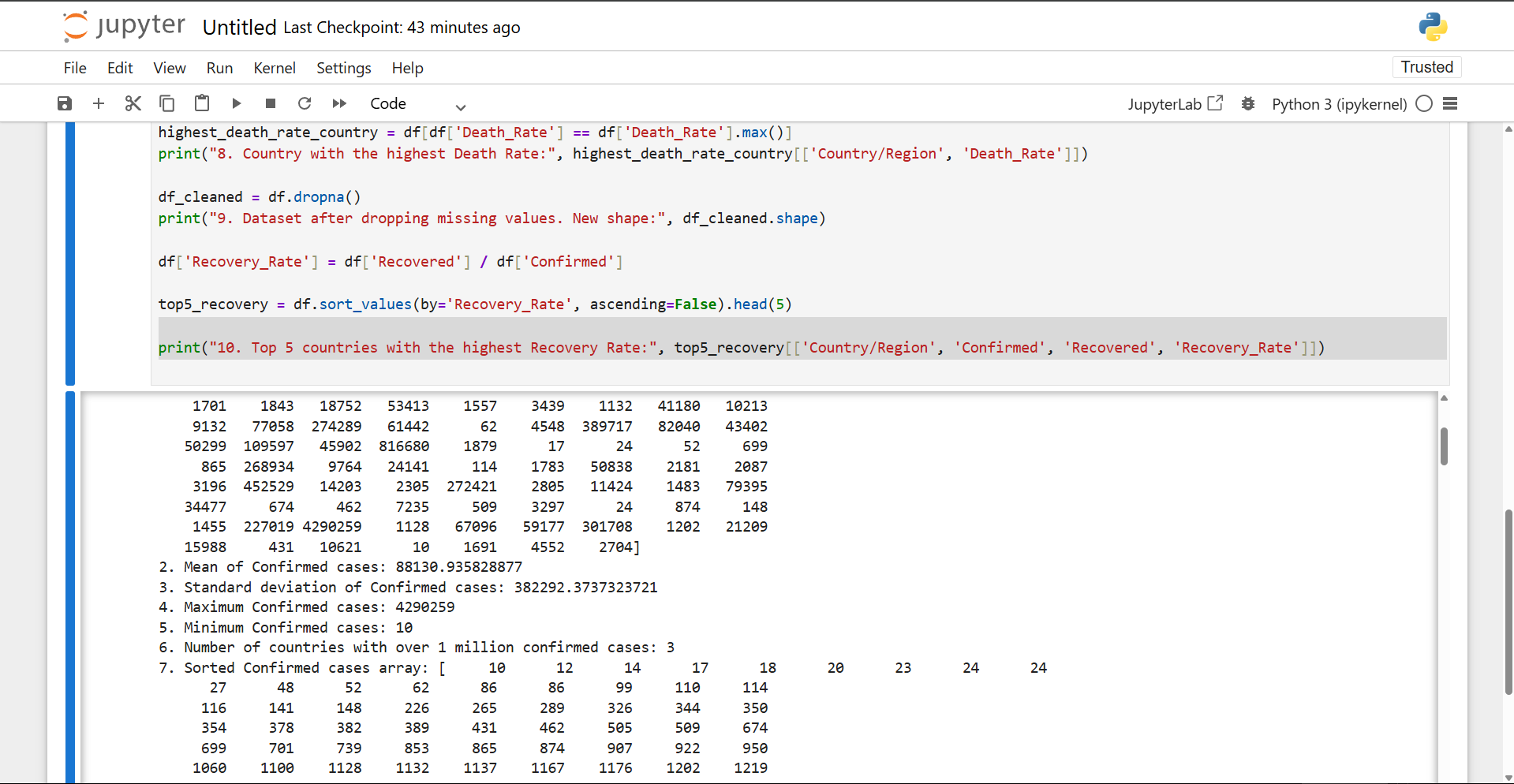
min\_confirmed = np.min(confirmed\_array)

print("5. Minimum Confirmed cases:", min\_confirmed)

Output: 

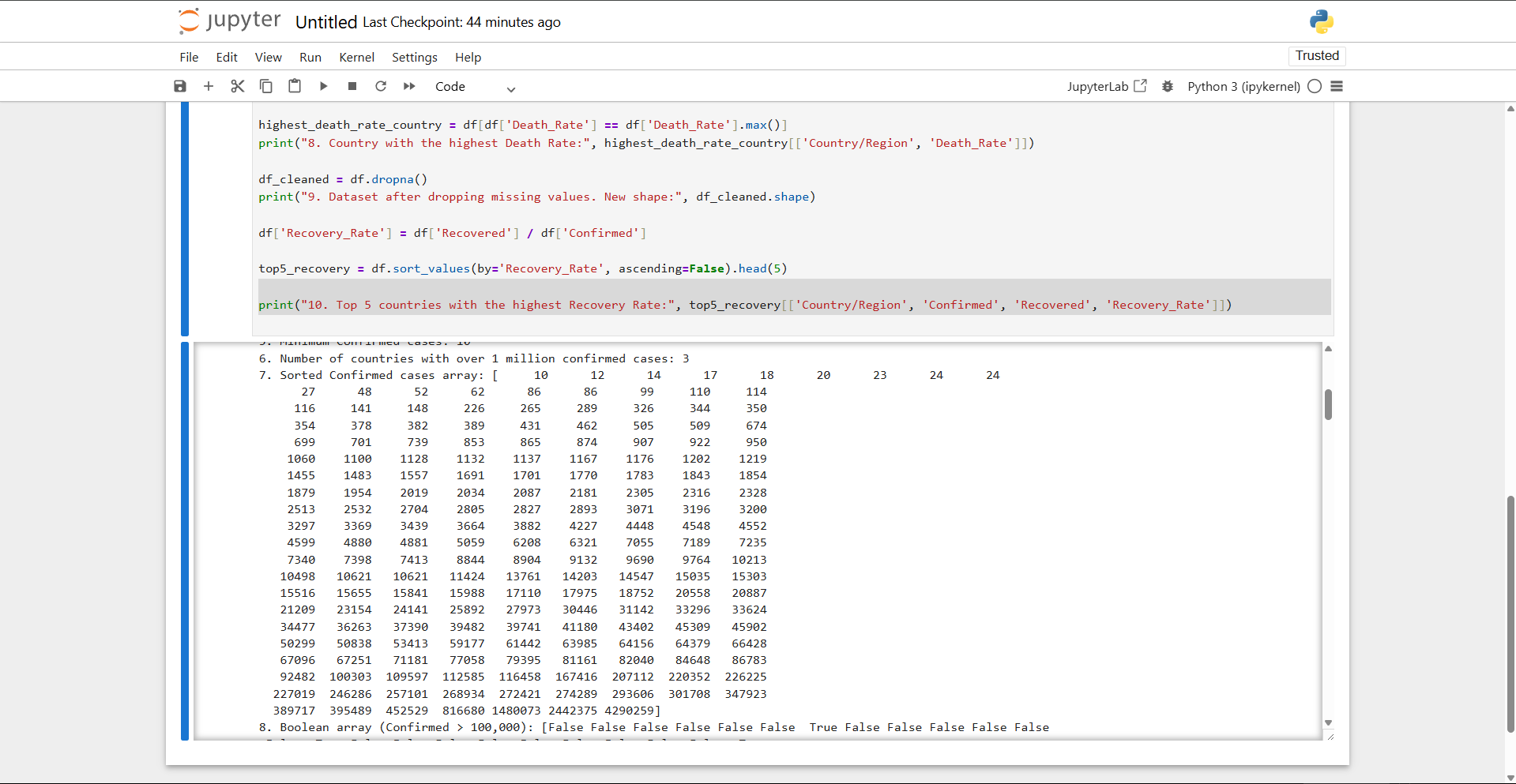
countries\_over\_million = np.sum(confirmed\_array > 1\_000\_000)

print("6. Number of countries with over 1 million confirmed cases:", countries\_over\_million)

Output: 

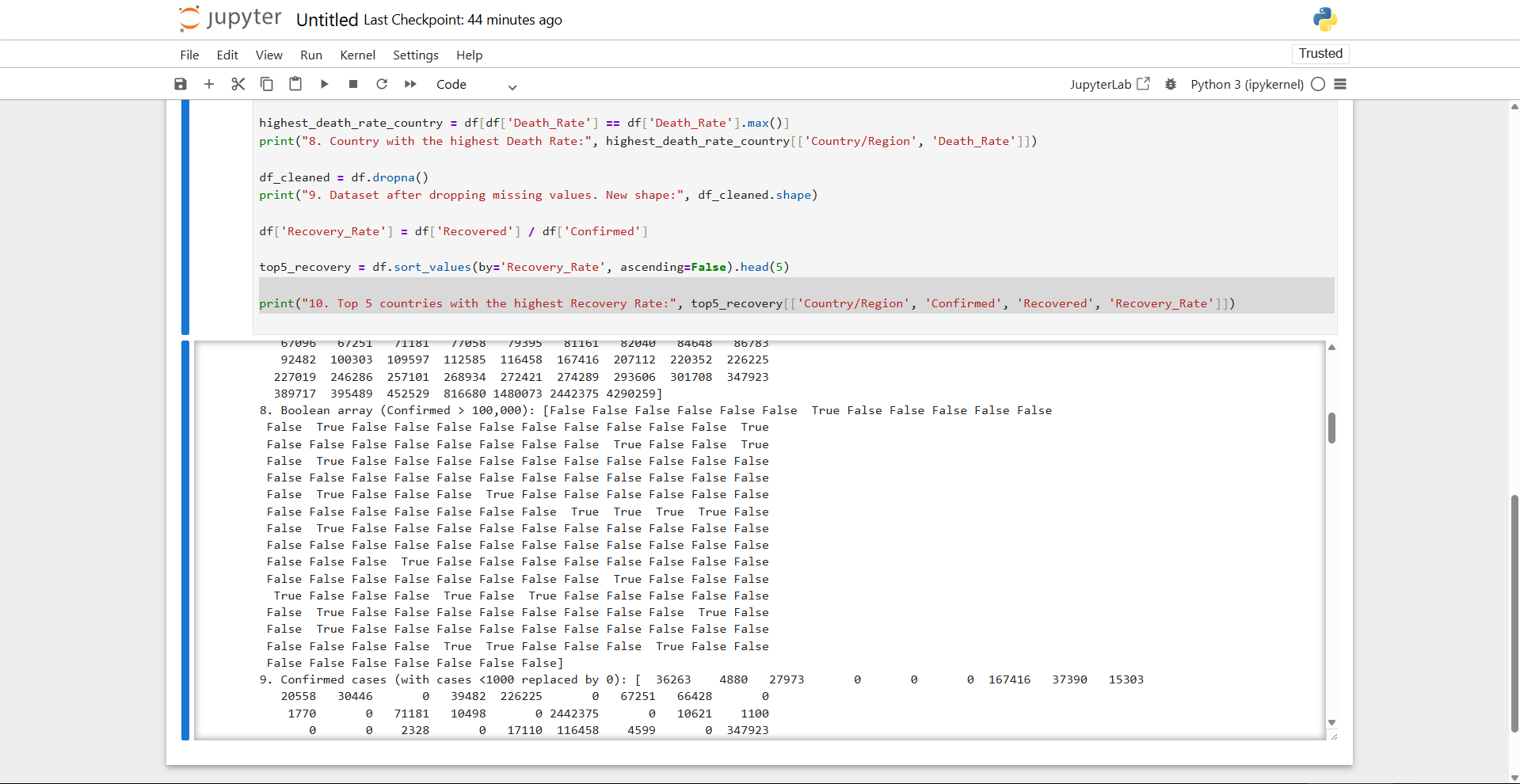
sorted\_confirmed = np.sort(confirmed\_array)

print("7. Sorted Confirmed cases array:", sorted\_confirmed)

Output: 

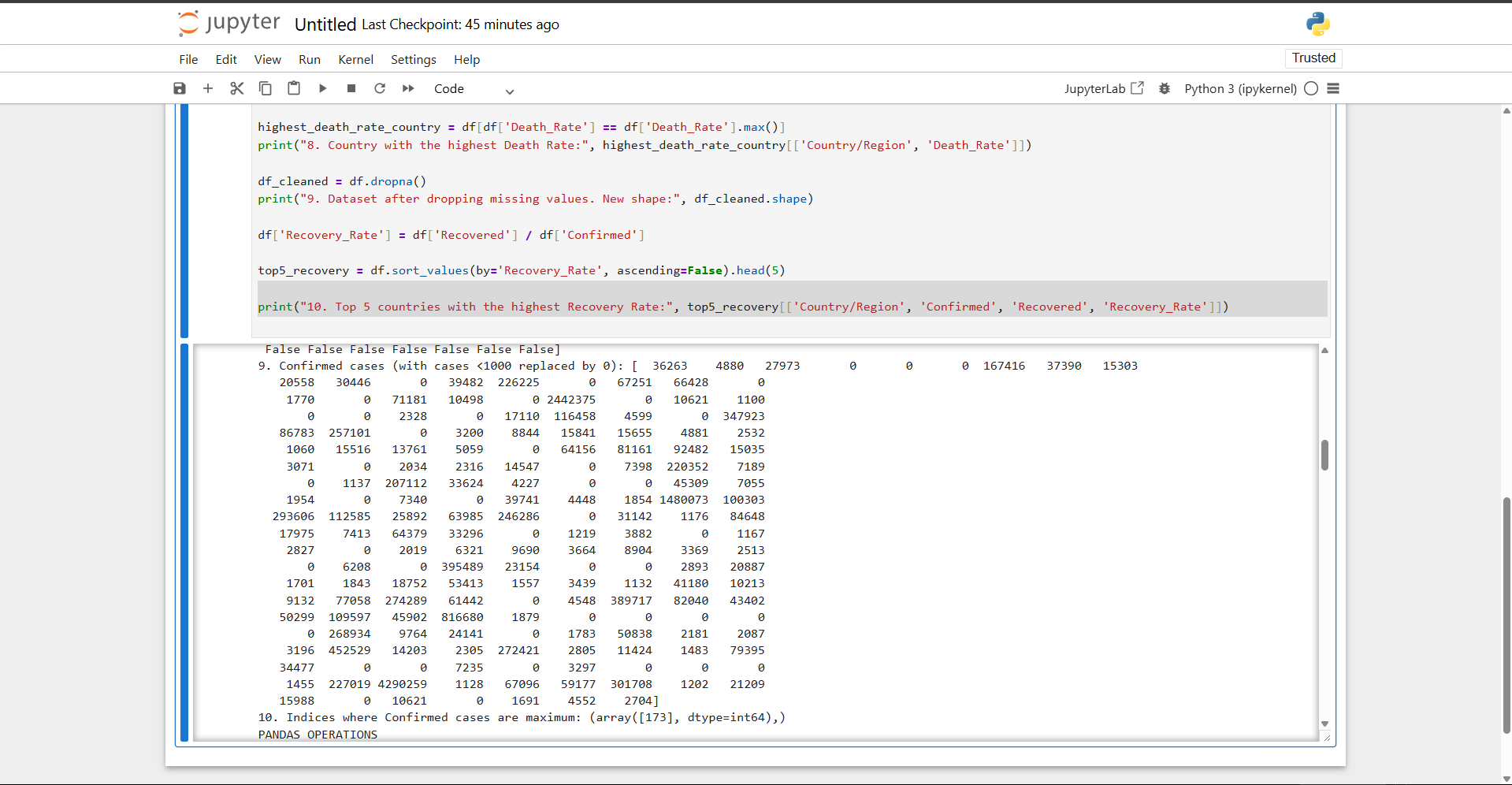
high\_confirmed\_bool = confirmed\_array > 100\_000

print("8. Boolean array (Confirmed > 100,000):", high\_confirmed\_bool)

Output: 

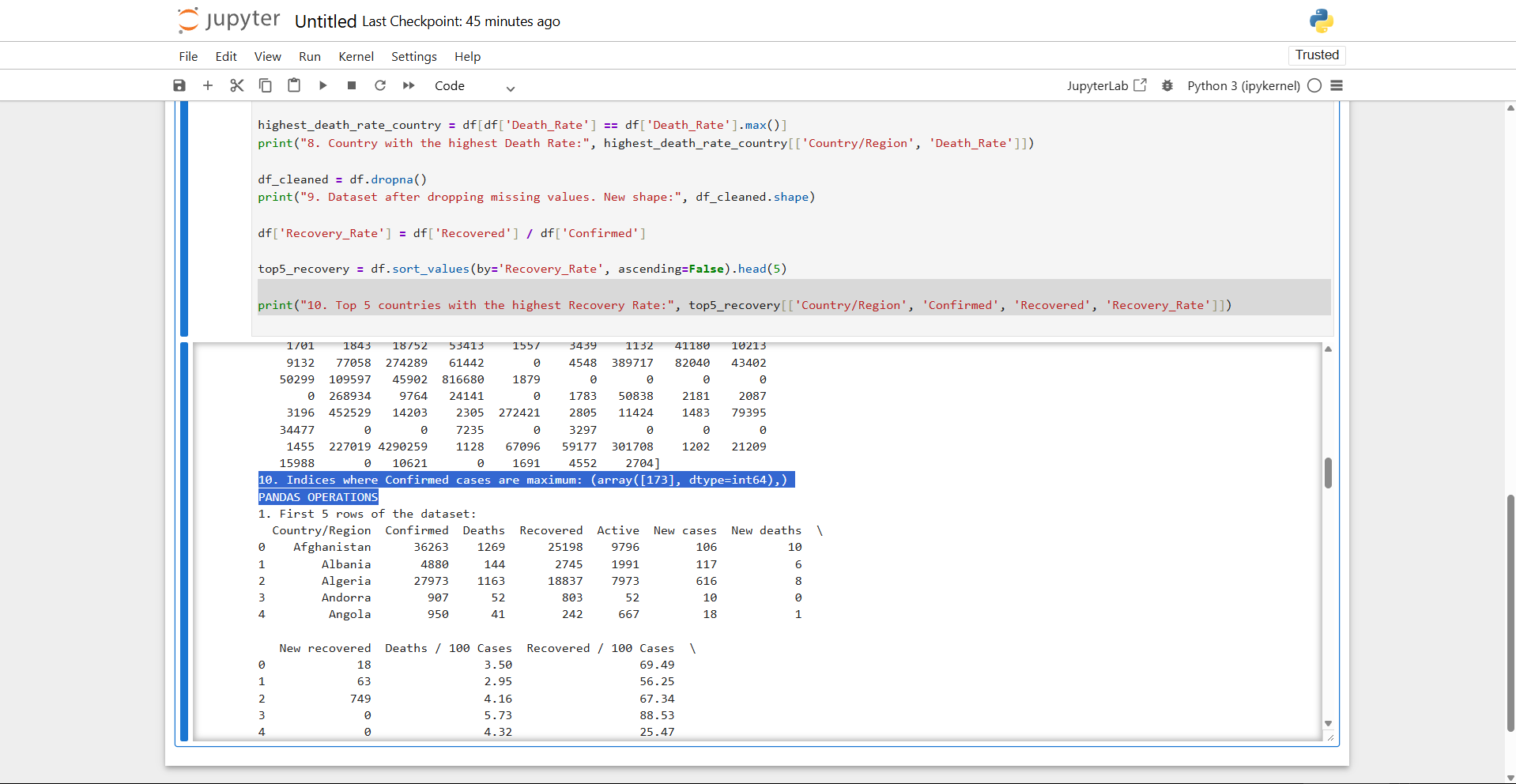
confirmed\_replaced = np.where(confirmed\_array < 1000, 0, confirmed\_array)

print("9. Confirmed cases (with cases <1000 replaced by 0):", confirmed\_replaced)

Output: 

max\_confirmed\_indices = np.where(confirmed\_array == np.max(confirmed\_array))

print("10. Indices where Confirmed cases are maximum:", max\_confirmed\_indices)

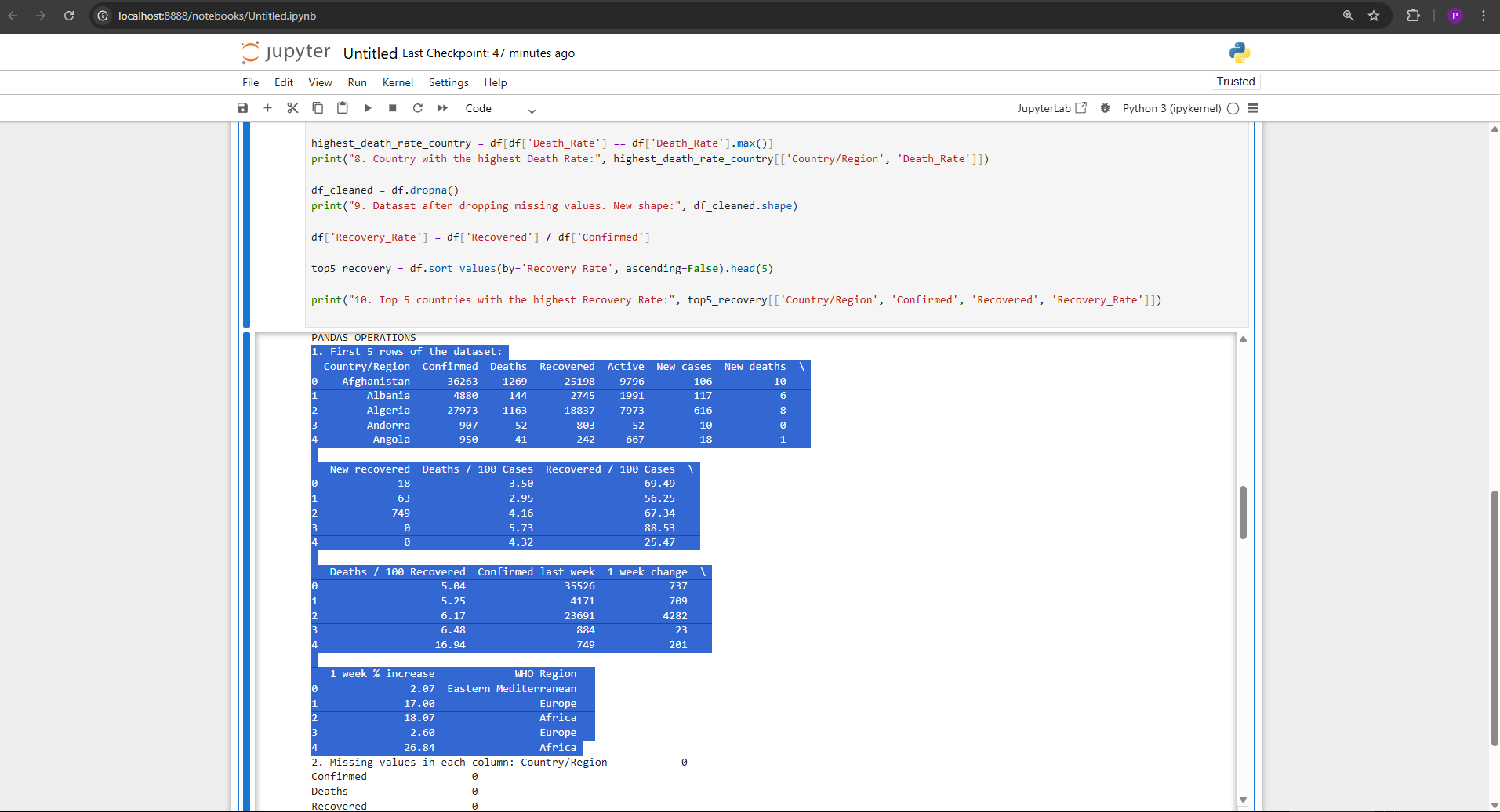
Output: 

#Highlighted Portion includes output for 10th and the starting for pandas function

print("PANDAS OPERATIONS")

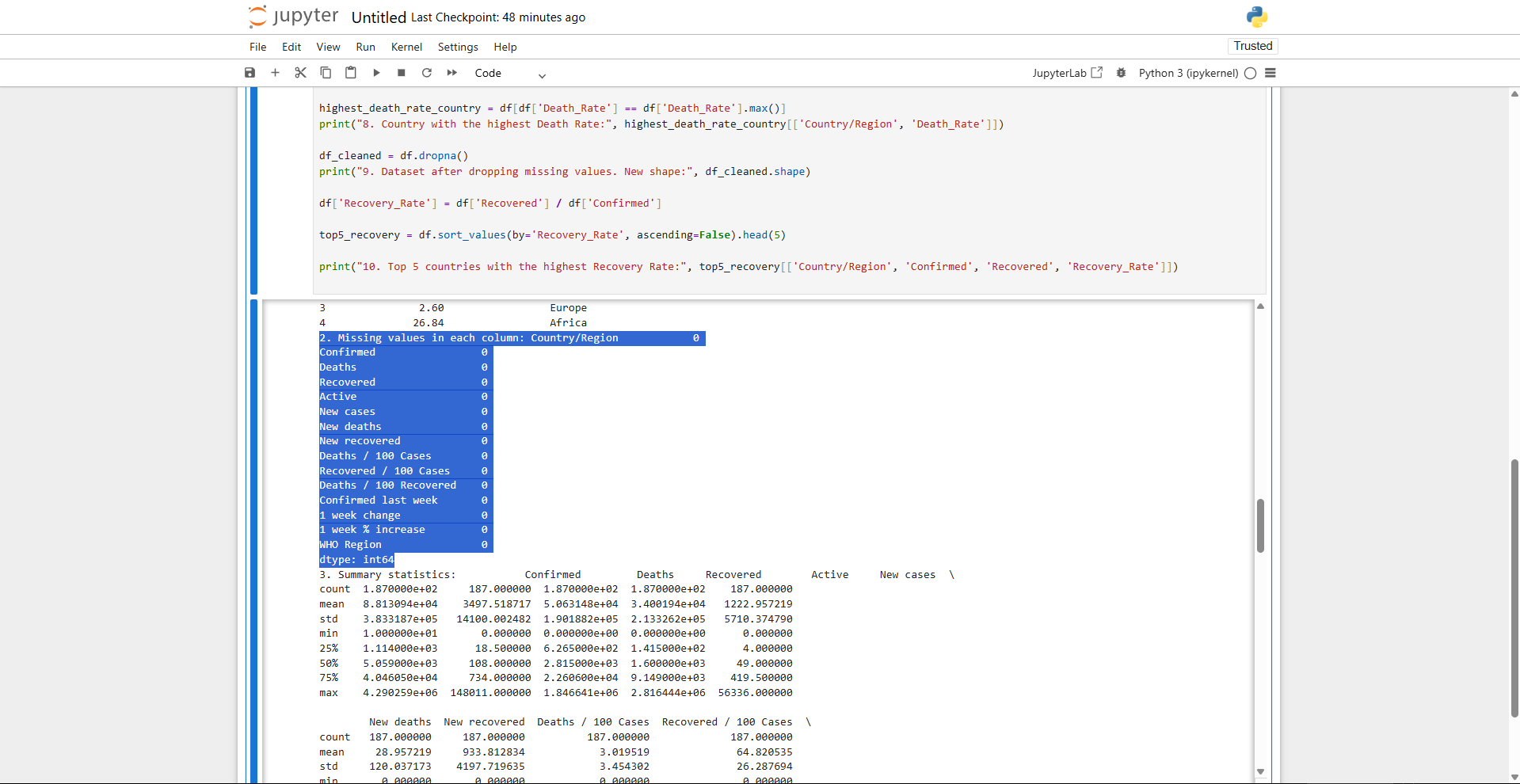
print("1. First 5 rows of the dataset:")

print(df.head())

output: 

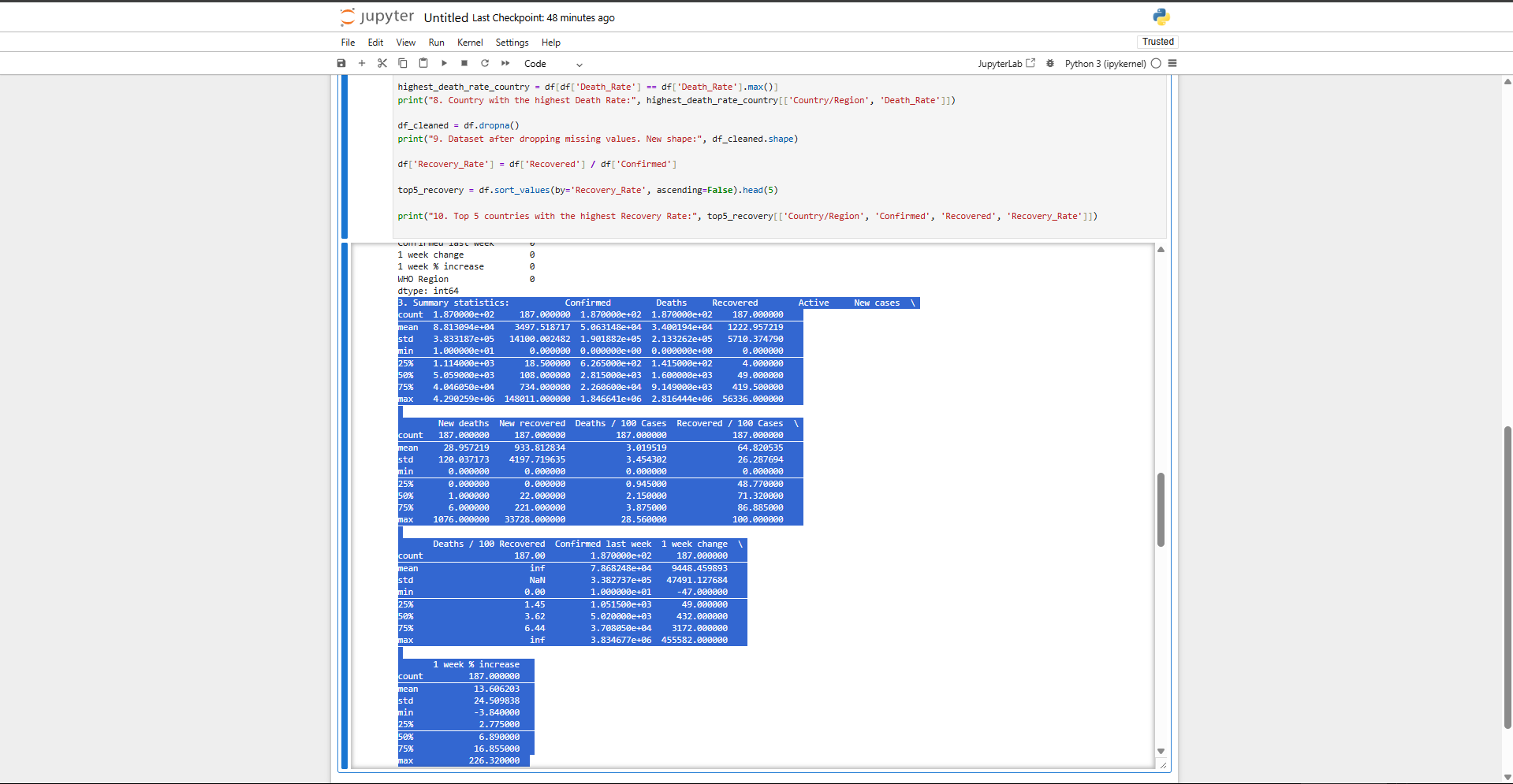
missing\_values = df.isnull().sum()

print("2. Missing values in each column:", missing\_values)

output: 

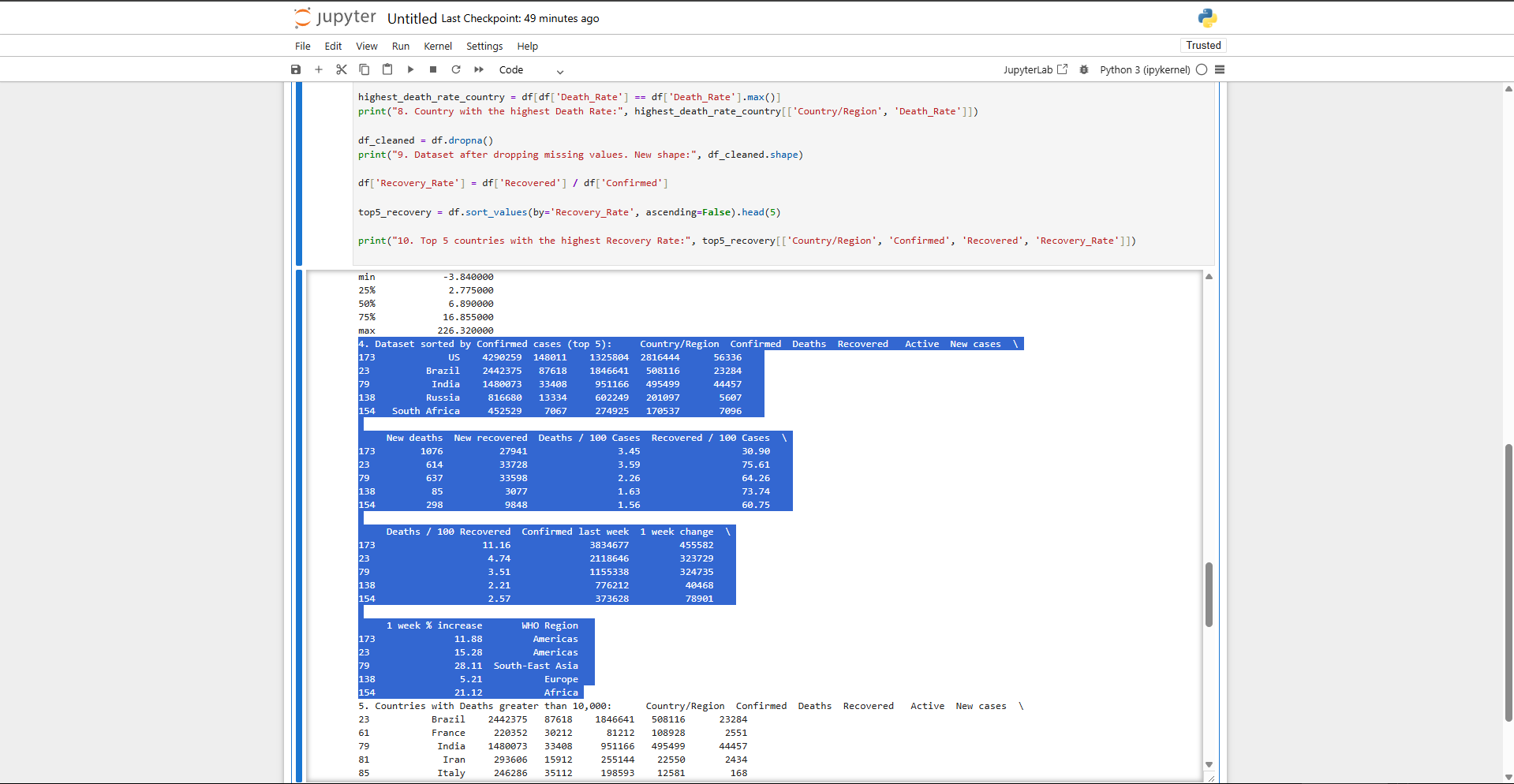
summary\_stats = df.describe()

print("3. Summary statistics:", summary\_stats)

Output: 

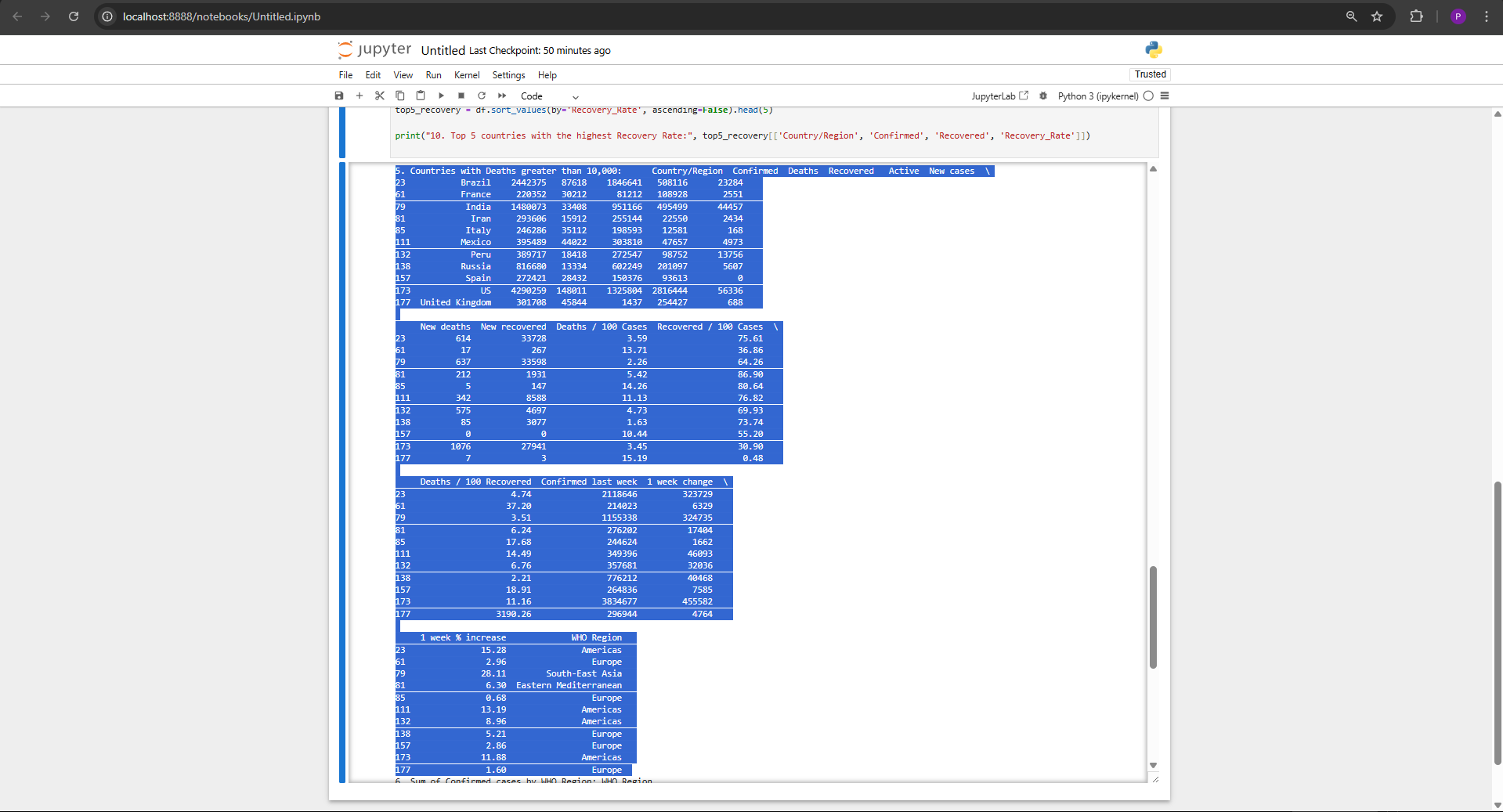
sorted\_df\_confirmed = df.sort\_values(by='Confirmed', ascending=False)

print("4. Dataset sorted by Confirmed cases (top 5):", sorted\_df\_confirmed.head())

Output: 

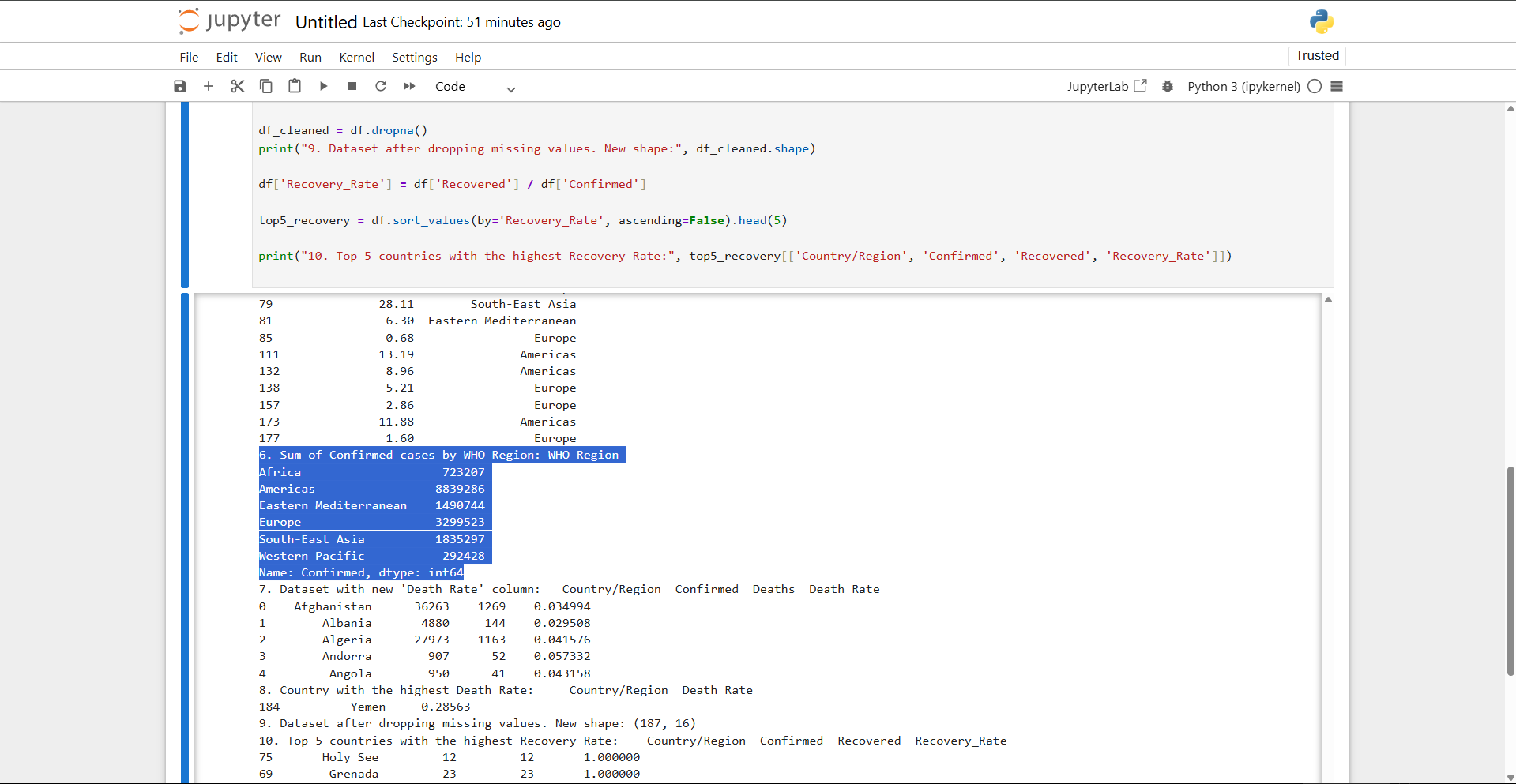
high\_deaths\_df = df[df['Deaths'] > 10000]

print("5. Countries with Deaths greater than 10,000:", high\_deaths\_df)

Output: 

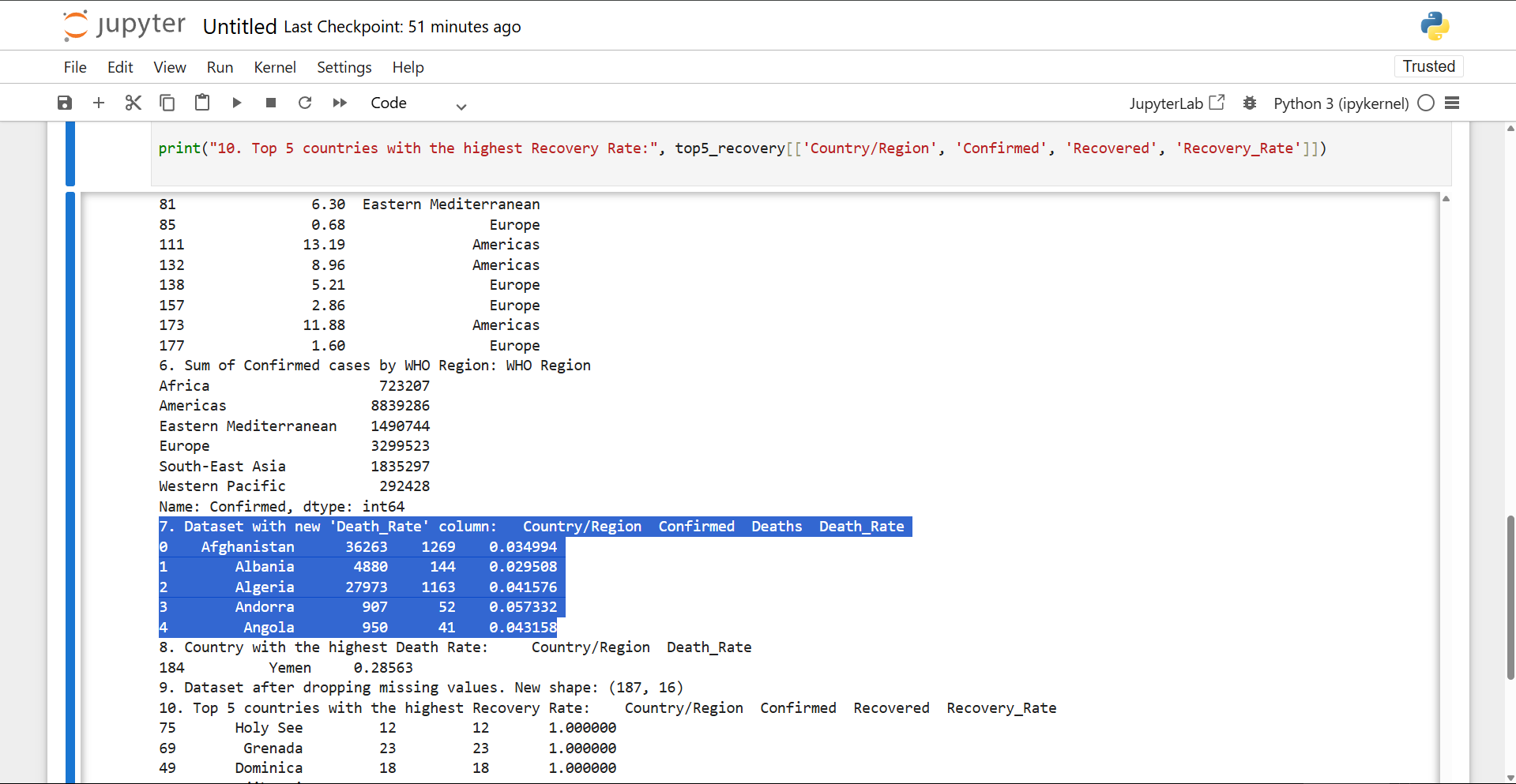
confirmed\_by\_region = df.groupby('WHO Region')['Confirmed'].sum()

print("6. Sum of Confirmed cases by WHO Region:", confirmed\_by\_region)

Output: 

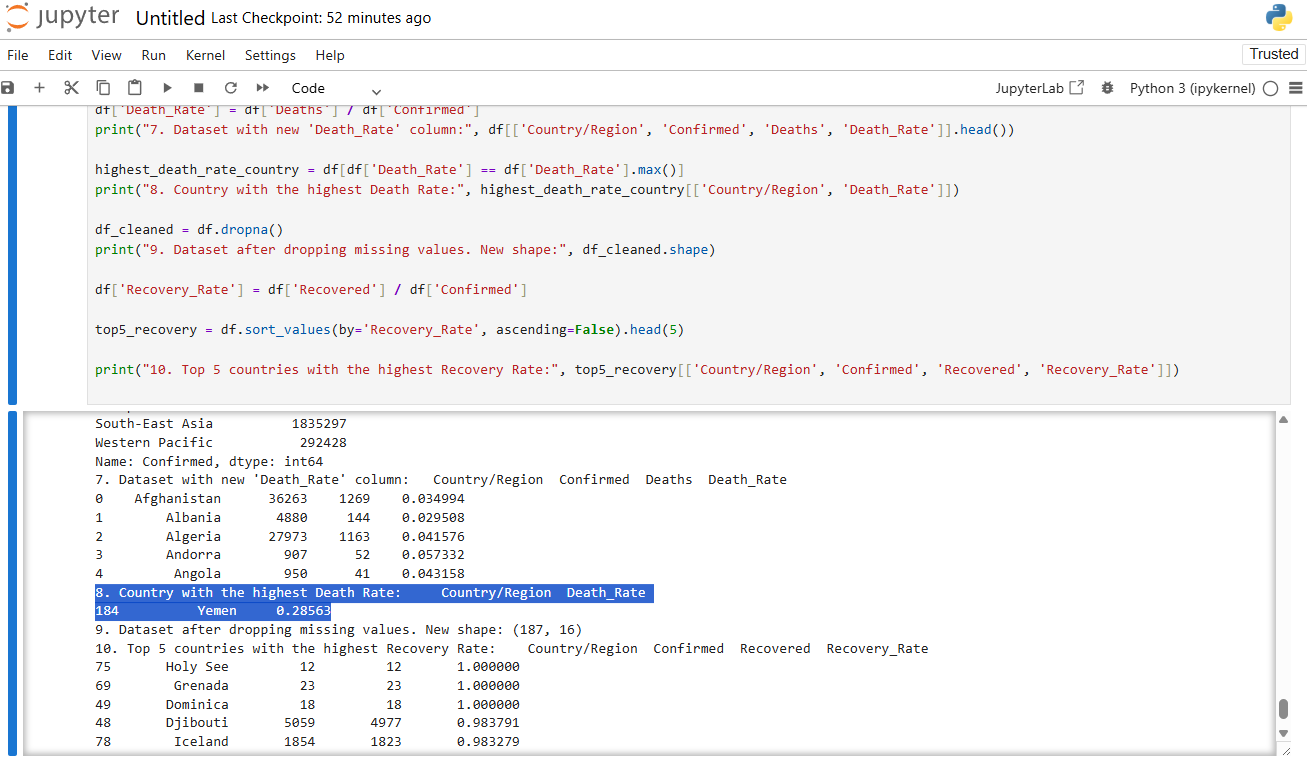
df['Death\_Rate'] = df['Deaths'] / df['Confirmed']

print("7. Dataset with new 'Death\_Rate' column:", df[['Country/Region', 'Confirmed', 'Deaths', 'Death\_Rate']].head())

Output: 

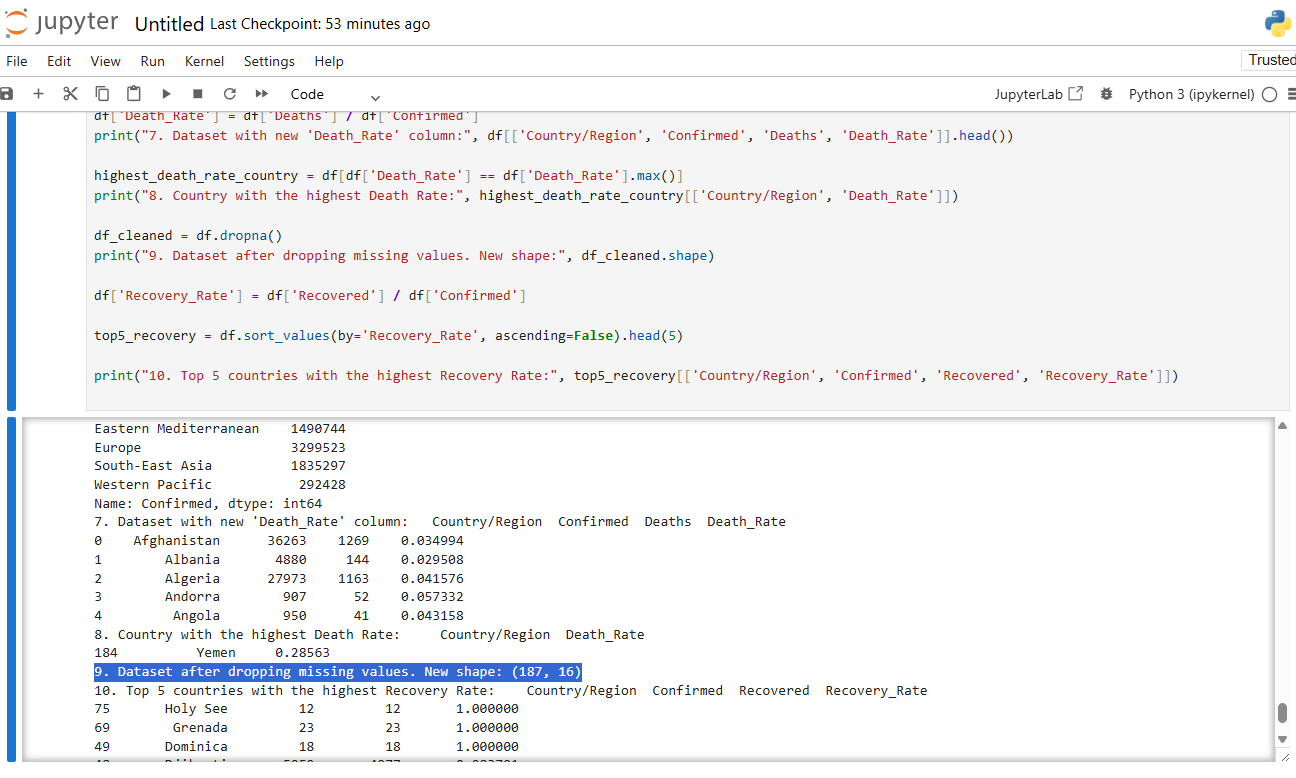
highest\_death\_rate\_country = df[df['Death\_Rate'] == df['Death\_Rate'].max()]

print("8. Country with the highest Death Rate:", highest\_death\_rate\_country[['Country/Region', 'Death\_Rate']])

Output: 

df\_cleaned = df.dropna()

print("9. Dataset after dropping missing values. New shape:", df\_cleaned.shape)

Output: 

df['Recovery\_Rate'] = df['Recovered'] / df['Confirmed']

top5\_recovery = df.sort\_values(by='Recovery\_Rate', ascending=False).head(5)

print("10. Top 5 countries with the highest Recovery Rate:", top5\_recovery[['Country/Region', 'Confirmed', 'Recovered', 'Recovery\_Rate']])

Output: 