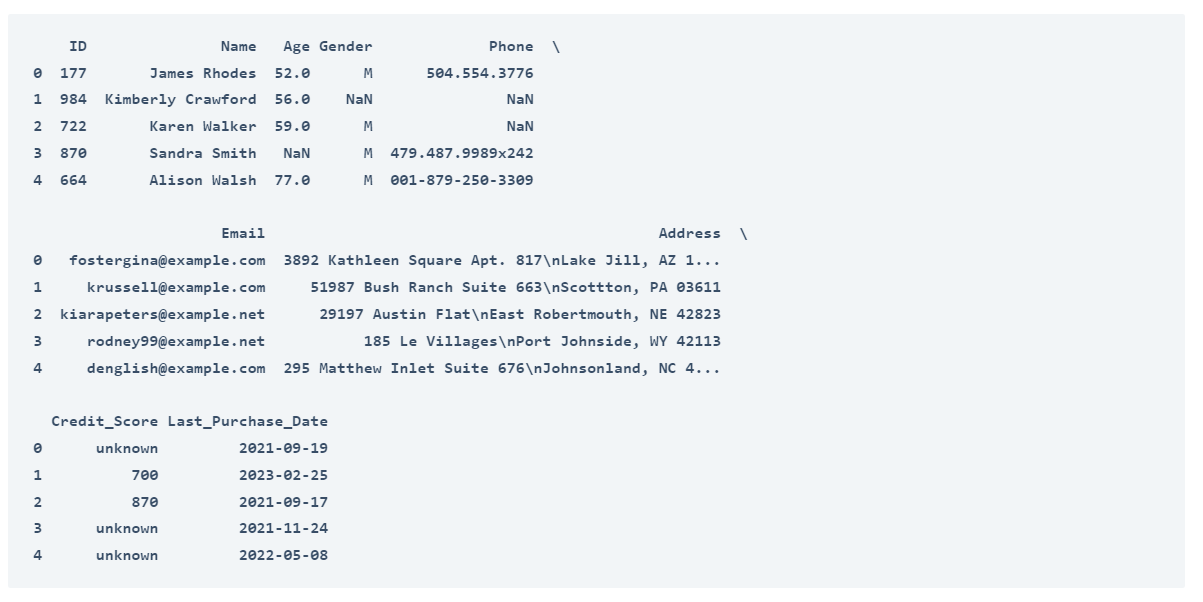
import pandas as pd

data = pd.read\_csv("sample\_data.csv")

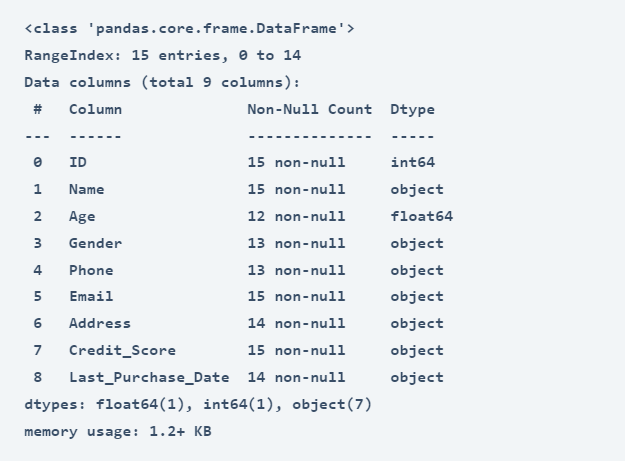
print(data.head())

#output



data.info()

#output



#cleaning the age column

# Convert 'Age' column to integer type

data['Age'] = data['Age'].astype('Int64')

median\_age = data['Age'].median()

median\_age = round(median\_age)

# Fill missing ages with the median value

data['Age'].fillna(median\_age, inplace=True)

# Handling missing values in 'Gender', 'Phone', and 'Address'

data['Gender'].fillna('Unknown', inplace=True)

data['Phone'].fillna('Unknown', inplace=True)

data['Address'].fillna('Unknown', inplace=True)

# Handling missing values in 'Last\_Purchase\_Date'

most\_recent\_date = pd.to\_datetime(data['Last\_Purchase\_Date']).max()

data['Last\_Purchase\_Date'].fillna(most\_recent\_date, inplace=True)

# Convert 'Credit\_Score' to numeric type and replace 'unknown' with NaN

data['Credit\_Score'] = pd.to\_numeric(data['Credit\_Score'],

errors='coerce')

# Calculate the mean of the available credit scores

mean\_credit\_score = data['Credit\_Score'].mean()

# Fill missing credit scores with the mean value

data['Credit\_Score'].fillna(mean\_credit\_score, inplace=True)

#droping duplicate rows from the data

data.drop\_duplicates(inplace=True)

print(data.head())

