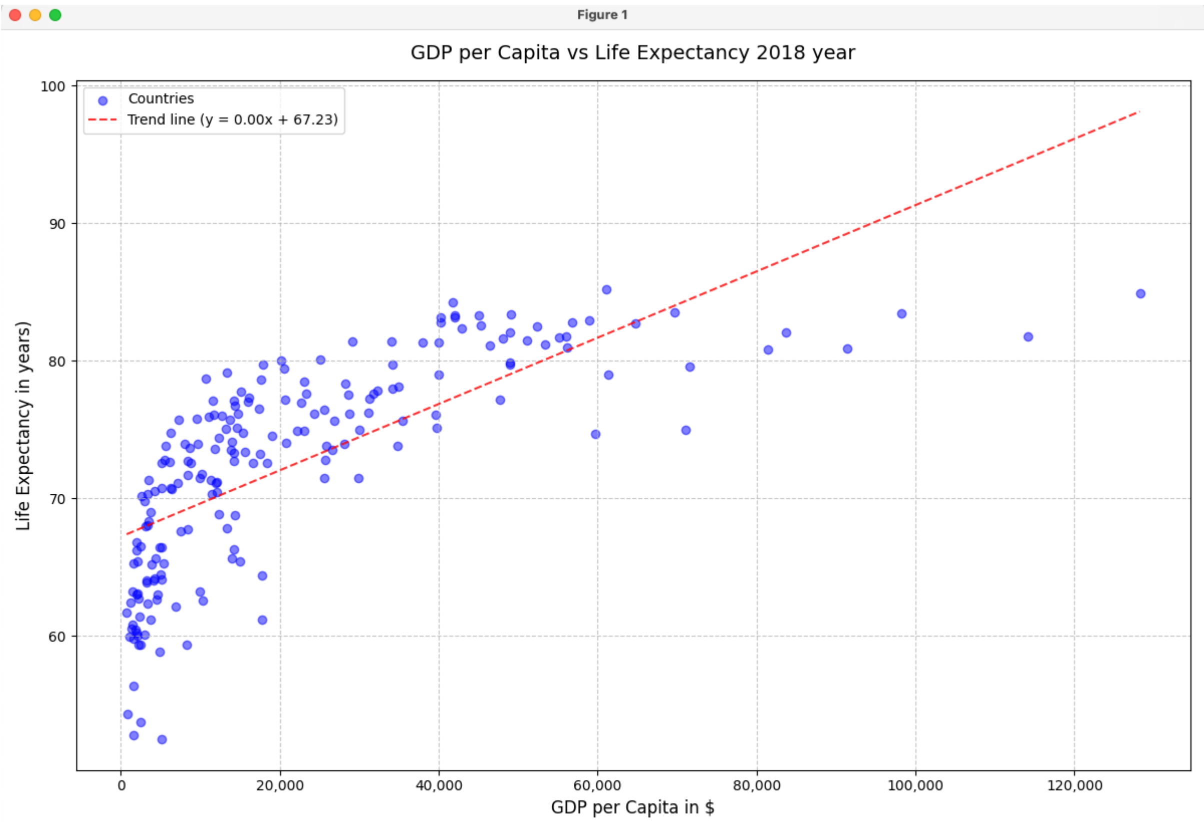
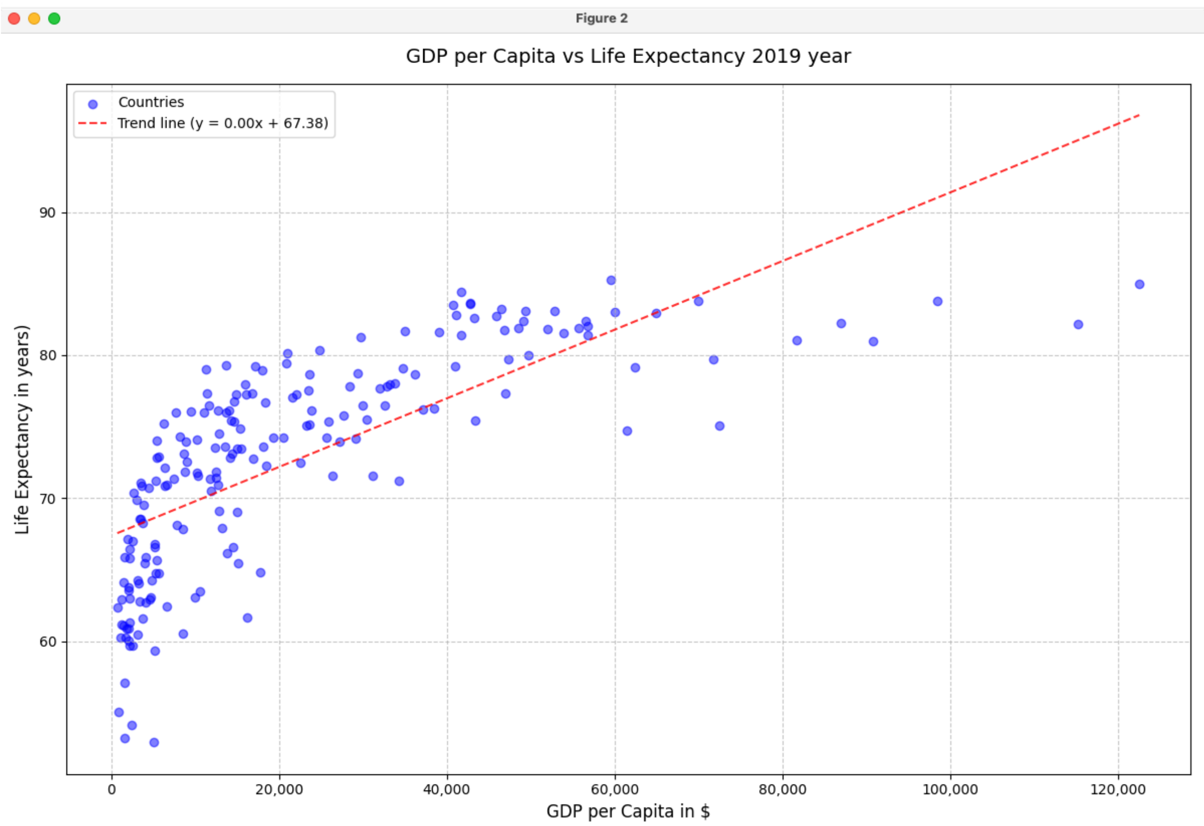
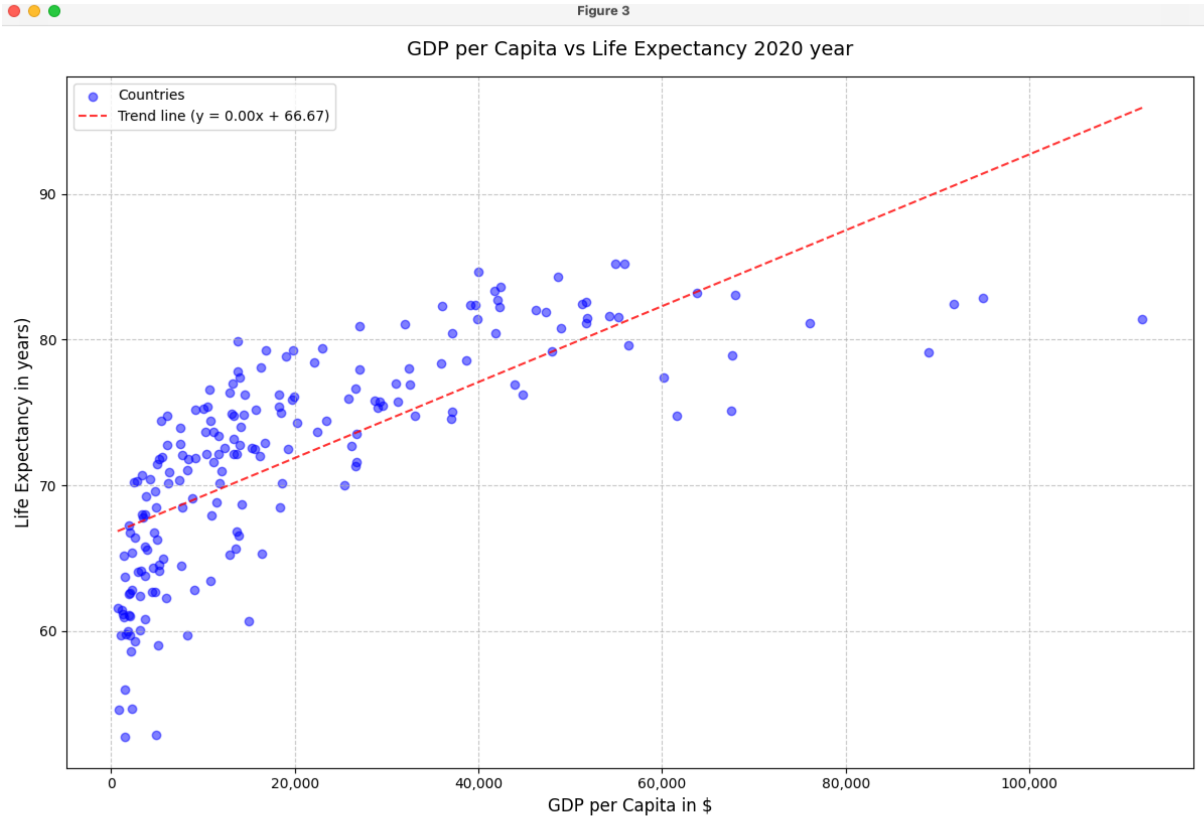
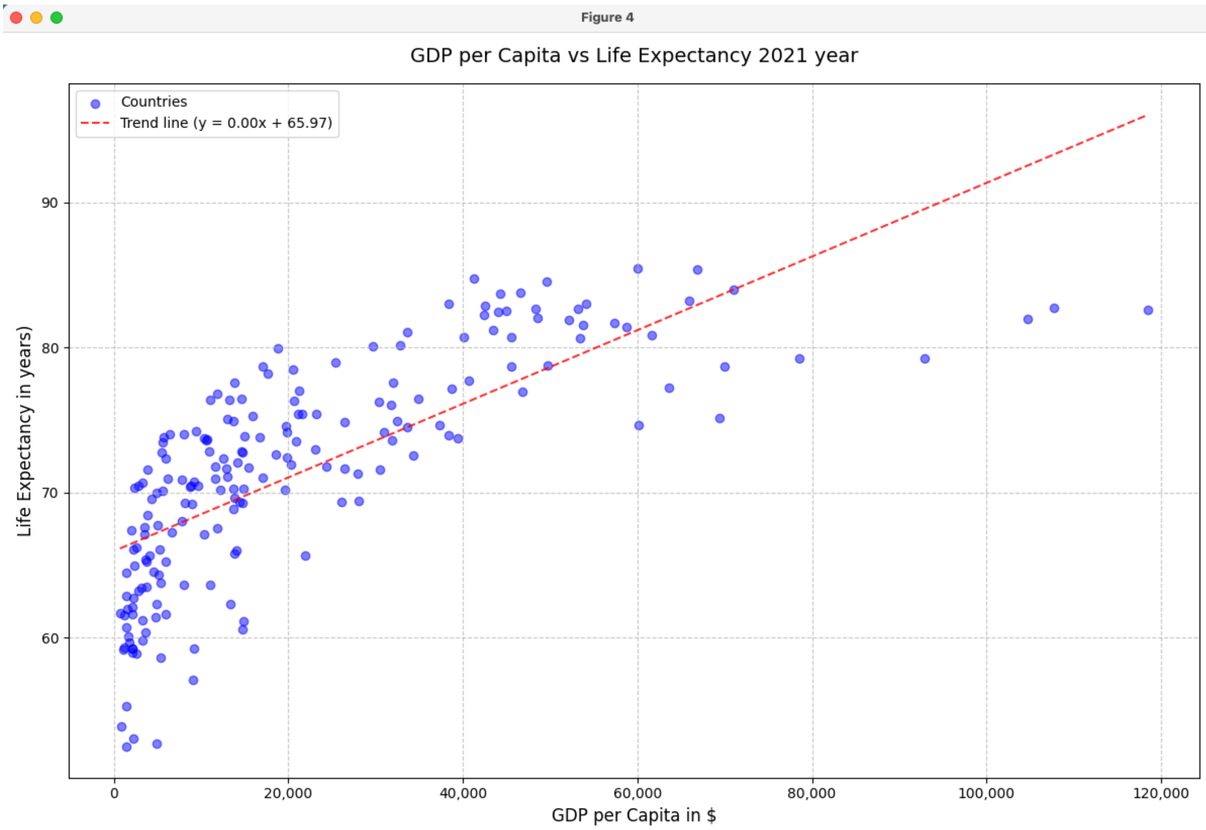
Scatter plot of GDP per capita vs Life Expectancy from 2018 to 2021









Sources of the data which is presented above can be found in the following links:
GDP Per Capita: https://ourworldindata.org/grapher/gdp-per-capita-worldbank
Life Expectancy: https://ourworldindata.org/grapher/life-expectancy

How the data is being plotted

1. Intersect the two data sets n ENTITY, CODE and YEAR
    - This ensures that the merged data set doesn't any corrupted rows
      For example:
      If the data is not merged on YEAR there migh be a row containing GDP per capita information for
      a certain year, but not containing the corresponding Life Expectancy age
2. Drop any rows that have NA data in some of the columns (Not available)
    - That removes any aggregated information like 'middle income', 'higher middle income', etc.
3. Scatter the GDP per capita values on X axis and the Life Expectancy on Y axis
4. Calculate (X,Y) coordinates which are ploted as a red trend line
    - For the X values I'm using the "linspace" method from numpy to generate 100 evenly spaced values between the minimum and maximum values of GDP Per Capita
    - For the Y coordinates I'm using "polyfit" and "pol1d" method from numpy

Countries that have higher than one standard deviation above the mean for 2021

```
Life expectancy threshold higher than 1 SD above mean:
        Entity  Life_Expectancy
     Australia          84.5265
       Austria          81.5797
       Belgium          81.8787
       Bermuda          79.2801
        Canada          82.6565
        Cyprus          81.2033
       Denmark          81.3753
       Finland          82.0381
        France          82.4988
       Germany          80.6301
        Greece          80.1106
     Hong Kong          85.4734
       Iceland          82.6782
       Ireland          81.9976
        Israel          82.2550
         Italy          82.8502
         Japan          84.7839
    Luxembourg          82.6287
         Macao          85.3994
      Maldives          79.9182
         Malta          83.7769
   Netherlands          81.6873
   New Zealand          82.4513
        Norway          83.2339
      Portugal          81.0443
   Puerto Rico          80.1618
         Qatar          79.2716
    San Marino          80.8786
     Singapore          82.7545
      Slovenia          80.6904
   South Korea          83.6978
         Spain          83.0100
        Sweden          82.9833
   Switzerland          83.9872
United Kingdom          80.7422
```

Contries with high Life Expectancy but low GDP per capita

How high/low thresholds are being calculated

1. Low GDP per capita threshold formula **– 60% of median**
   In unequal distributions, the mean is likely to be influenced by high values, so it does not reflect most countries. The median is not affected by a few very high values.
   More information can found in [UK Gov website]
2. High GDP per capita threshold formula – **150% of median**
   Setting the high threshold for GDP per capita at 150% of the median ensures that only countries with GDP per capita significantly above the middle range are considered high. Also, as stated for the low GDP formula the median is not influenced by extreme outliers.
3. Low Life Expectancy Age formula – **The bottom 35% of the values**
   This percentile approach is commonly used in government and health policy analysis. The bottom 35% might include a lot of countries. However, since health care is crucial topic for each country, it's good to identify Life Expectancy as 'low' even if it's slightly lower than the average
4. High Life Expectancy Age formula – **The top 15% of the values**
   A threshold of the top 15% for high life expectancy is designed to distinguish countries where life expectancy is substantially above the norm. This method ensures that only the countries with the longest life expectancies are included.

Using the mentioned low/high thresholds there aren't any countries with High Life Expectancy but Low GDP per capita in the data set for 2018, 2019, 2020, 2021.
To get this result I took the intersection of the low life expectancy countries and the high GDP countries.

Does every strong economy (normally indicated by GDP per capita) have high life expectancy?

Most of the times high GDP correlates with high Life Expectancy. The correlation coefficient between GDP and Life Expectancy of this dataset is **0,738** which indicates as strong relationship between the two metrics. This relationship aligns with economic and health research, where countries with higher GDP per capita often have better healthcare systems, higher standards of living, and thus, longer life expectancies.

However, there are a few outliers to this relationship.

- 2020

```
Countries with high GDP per capita and low life expectancy:
    Entity Code  Year  Life_Expectancy  GDP_per_capita
Kazakhstan  KAZ  2020          70.0304       25361.508
```

- 2021

```
Countries with high GDP per capita and low life expectancy:
    Entity Code  Year  Life_Expectancy  GDP_per_capita
    Guyana  GUY  2021          65.6734       21925.223
Kazakhstan  KAZ  2021          69.3622       26110.530
```

For 2018 and 2019 there aren't any outliers corresponding the High/Low thresholds set above. But 2020 and 2021 indicate that while GDP per capita and life expectancy are related, this doesn't confirm that one directly causes the other. Other factors like healthcare access, education, and government policies also play crucial roles.

What would happen if you used GDP per capita as an indicator of strong economy?

GDP per capita is currently being used as a one of the measures of a strong economy. However, there are other metrics which should be taken into consideration when trying to identify a strong economy.

- **Gross National Product (GNP):** Measures the total value of goods and services produced by a country's residents, regardless of their location.
- **Inflation Rate:** Measures the rate at which prices for goods and services increase over time. Low and stable inflation is generally seen as positive.
- **Unemployment Rate:** Measures the percentage of the labor force that is unemployed and actively seeking work. A low unemployment rate indicates a strong job market.
- **Interest Rates:** The cost of borrowing money. Central banks often adjust interest rates to influence economic activity.
- **Productivity:** Measures how efficiently resources are used to produce goods and services.

In conclusion, high GDP per capita often indicates a strong economy, but for certain there are outliers. For example, if there is hyperinflation and high unemployment rate a high GDP doesn't mean the economy is healthy. As stated in the previous topic high life expectancy doesn't always correlate with high GDP per capita and vice-versa. The same applies to correlation between high GDP per capita and a strong economy.

Python code

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

LIFE_EXPENTANCY_CSV = 'life-expectancy.csv'
GDP_PER_CAPITA_CSV = 'gdp-per-capita-worldbank.csv'
LIFE_EXPENTANCY = 'Life_Expectancy'
GDP_PER_CAPITA = 'GDP_per_capita'
ENTITY = 'Entity'
CODE = 'Code'
YEAR = 'Year'
YEARS_TO_ANALYZE = [1990, 1998, 2008, 2018, 2019, 2020, 2021]

def init_data():
    life_expectancy = pd.read_csv(LIFE_EXPENTANCY_CSV)
    gdp = pd.read_csv(GDP_PER_CAPITA_CSV)

    life_expectancy = life_expectancy.rename(columns={
        'Period life expectancy at birth - Sex: all - Age: 0': LIFE_EXPENTANCY
    })
    gdp = gdp.rename(columns={
        'GDP per capita, PPP (constant 2017 international $)': GDP_PER_CAPITA
    })

    return pd.merge(life_expectancy, gdp,
                    on=[ENTITY, CODE, YEAR],
                    how='inner')


def filter_data_by_year(data, year):
    filtered_data = data[data[YEAR] == year]
    return filtered_data.dropna()

def format_func(value, tick_number):
    return f'{int(value):,}'

def plot_data_by_year(data, year):
    plt.figure(figsize=(12, 8))
    plt.scatter(data[GDP_PER_CAPITA],
            data[LIFE_EXPENTANCY],
            alpha=0.5,
            c='blue',
            label='Countries')
    z = np.polyfit(data[GDP_PER_CAPITA], data[LIFE_EXPENTANCY], 1)
    p = np.poly1d(z)
    x_trend = np.linspace(data[GDP_PER_CAPITA].min(),
                    data[GDP_PER_CAPITA].max(),
                    100)
```

```
    plt.plot(x_trend, p(x_trend), "r--", alpha=0.8,
            label=f'Trend line (y = {z[0]:.2f}x + {z[1]:.2f})')
    plt.title(f'GDP per Capita vs Life Expectancy {year} year', fontsize=14,
pad=15)
    plt.xlabel('GDP per Capita in $', fontsize=12)
    plt.ylabel('Life Expectancy in years)', fontsize=12)

    plt.gca().xaxis.set_major_formatter(plt.FuncFormatter(format_func))
    plt.grid(True, linestyle='--', alpha=0.7)
    plt.legend()
    plt.tight_layout()


def get_hight_expectancy_low_gdp_countries(data):
    high_life_exp_countries = get_high_end_countries_by_life_exp(data)
    low_gdp_countries = get_low_end_countries_by_gdp(data)

    return pd.merge(high_life_exp_countries, low_gdp_countries, how='inner',
on=[ENTITY, CODE, YEAR, GDP_PER_CAPITA, LIFE_EXPENTANCY])

def get_high_gdp_low_life_exp_countries(data):
    high_gdp_countries = get_high_end_countries_by_gdp(data)
    low_life_exp_countries = get_low_end_countries_by_life_exp(data)

    return pd.merge(high_gdp_countries, low_life_exp_countries, how='inner',
on=[ENTITY, CODE, YEAR, GDP_PER_CAPITA, LIFE_EXPENTANCY])


def get_high_end_countries_by_gdp(data):
    threshold = calculate_high_gdp_threshold(data[GDP_PER_CAPITA])
    high_life_exp_countries = data[data[GDP_PER_CAPITA] > threshold].copy()

    return
high_life_exp_countries.loc[high_life_exp_countries.groupby('Entity')[GDP_PER_CAPIT
A].idxmax()]

def get_low_end_countries_by_gdp(data):
    threshold = calculate_low_gdp_threshold(data[GDP_PER_CAPITA])
    low_life_exp_countries = data[data[GDP_PER_CAPITA] < threshold].copy()

    return
low_life_exp_countries.loc[low_life_exp_countries.groupby('Entity')[GDP_PER_CAPITA]
.idxmin()]

def get_high_end_countries_by_life_exp(data):
    threshold = calculate_high_life_exp_threshold(data[LIFE_EXPENTANCY])
    high_life_exp_countries = data[data[LIFE_EXPENTANCY] > threshold].copy()

    return
high_life_exp_countries.loc[high_life_exp_countries.groupby('Entity')[LIFE_EXPENTAN
CY].idxmax()]
```

```
def get_low_end_countries_by_life_exp(data):
    threshold = calculate_low_life_exp_threshold(data[LIFE_EXPENTANCY])
    low_life_exp_countries = data[data[LIFE_EXPENTANCY] < threshold].copy()

    return
low_life_exp_countries.loc[low_life_exp_countries.groupby('Entity')[LIFE_EXPENTANCY
].idxmin()]

def get_countries_with_higher_than_one_std(data, metric_key):
    threshold = calculate_threshold_higher_than_one_std(data[metric_key])
    return data[data[metric_key] > threshold]

def calculate_threshold_higher_than_one_std(values):
    return values.mean() + values.std()

def calculate_high_life_exp_threshold(values):
    return values.quantile(0.85)

def calculate_low_life_exp_threshold(values):
    return values.quantile(0.35)

def calculate_high_gdp_threshold(values):
    return values.median()*1.5

def calculate_low_gdp_threshold(values):
    return values.median()*0.6

def run_statistics_for(years):
    data = init_data()

    for year in years:
        filtered_data = filter_data_by_year(data, year)

        # b)
        higher_than_one_std_conties =
get_countries_with_higher_than_one_std(filtered_data, LIFE_EXPENTANCY)
        print(f'\nLife expectancy threshold higher than 1 SD above mean for
{year}:')
        print(higher_than_one_std_conties[[ENTITY,
LIFE_EXPENTANCY]].to_string(index=False))

        # c)
        print(f'\nCountries with high life expectancy and low GDP per capita for
{year}:')

print(get_hight_expectancy_low_gdp_countries(filtered_data).to_string(index=False))

        # d)
        high_gdp_low_life_exp_countries =
get_high_gdp_low_life_exp_countries(filtered_data)
```

```
        print(f'\nCountries with high GDP per capita and low life expectancy for
{year}:')
        print(high_gdp_low_life_exp_countries.to_string(index=False))
        correlation = np.corrcoef(filtered_data[GDP_PER_CAPITA],
                                filtered_data[LIFE_EXPENTANCY])[0,1]
        print("\nCorrelation coefficient between GDP per capita and Life
Expectancy:",
            round(correlation, 3))

        # a)
        plot_data_by_year(filtered_data, year)


    print('\nPlotting...')
    plt.show()

run_statistics_for(YEARS_TO_ANALYZE)
```


Author: Mihael Zlatev
Faculty Number: 1MI3400543