



PROJEKT ANALYSEWERKZEUGE

Daniel Braunstein

Mat.Nr.: 52000347

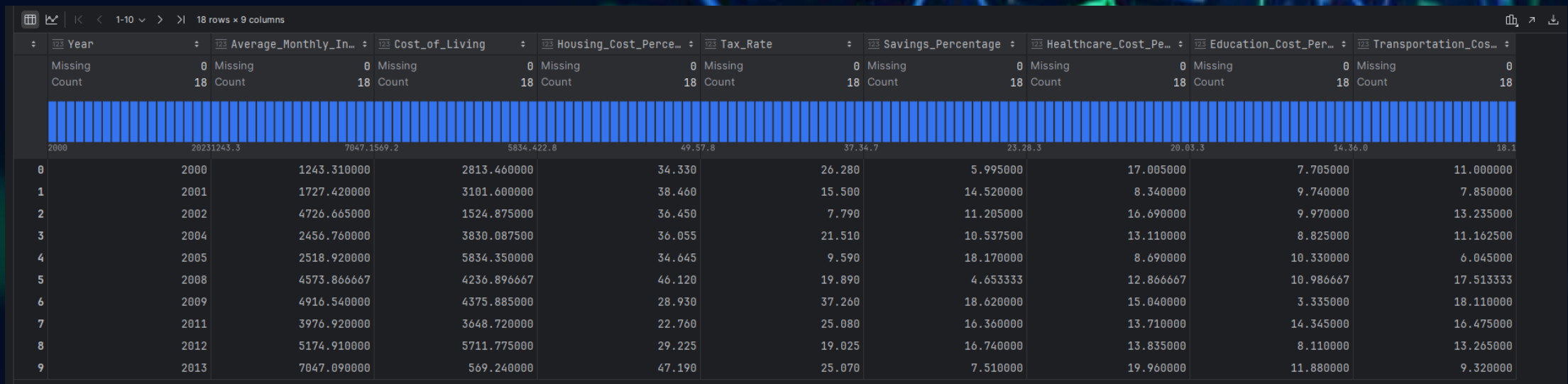
INHALT

- Datensatz
- Problemstellung
- Analyse
- Ergebnisse



DATENSATZ

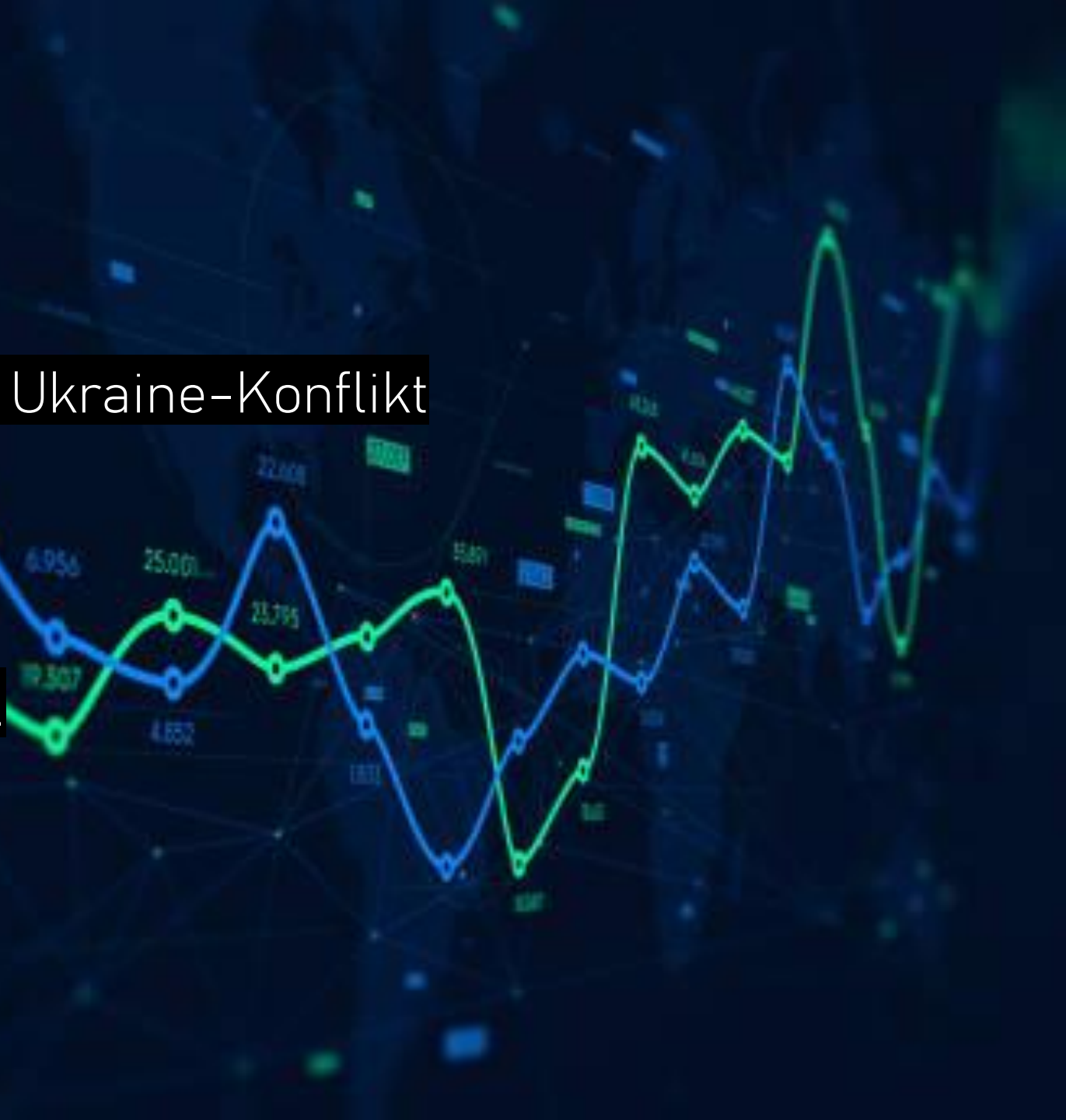
- Regional Cost of Living Analysis (von 2000 – 2023)
- <https://www.kaggle.com/datasets/heidarmirhajisadati/regional-cost-of-living-analysis>
- <https://www.kaggle.com/code/danielbraunstein/daniel-braunstein-fhwn-adi-ddi>



	Year	Average_Monthly_Income	Cost_of_Living	Housing_Cost_Percentage	Tax_Rate	Savings_Percentage	Healthcare_Cost_Percentage	Education_Cost_Percentage	Transportation_Cost_Percentage
Missing Count	0	Missing Count 18	0	Missing Count 18	0	Missing Count 18	0	Missing Count 18	0
	2000	1243.310000	2813.460000	34.330	26.280	5.995000	17.005000	7.705000	11.000000
1	2001	1727.420000	3101.600000	38.460	15.500	14.520000	8.340000	9.740000	7.850000
2	2002	4726.665000	1524.875000	36.450	7.790	11.205000	16.690000	9.970000	13.235000
3	2004	2456.760000	3830.087500	36.055	21.510	10.537500	13.110000	8.825000	11.162500
4	2005	2518.920000	5834.350000	34.645	9.590	18.170000	8.690000	10.330000	6.045000
5	2008	4573.866667	4236.896667	46.120	19.890	4.653333	12.866667	10.986667	17.513333
6	2009	4916.540000	4375.885000	28.930	37.260	18.620000	15.040000	3.335000	18.110000
7	2011	3976.920000	3648.720000	22.760	25.080	16.360000	13.710000	14.345000	16.475000
8	2012	5174.910000	5711.775000	29.225	19.025	16.740000	13.835000	8.110000	13.265000
9	2013	7047.090000	569.240000	47.190	25.070	7.510000	19.960000	11.880000	9.320000

PROBLEM

- Finanzkrise durch Corona und Ukraine-Konflikt
- Zinssteigerung der Banken
- Erhöhung der Lebenskosten
- Insolvenzen, Arbeitslosigkeit,...



ANALYSEANSATZ

- Wie verhalten sich Einnahmen und Ausgabe?
- Wie wird sich dies in Zukunft verhalten?
- Gibt es Ähnlichkeiten?
- Was kann aus den Korrelationen abgelesen werden?

ANALYSEANSATZ

- Diagramm AMI & CoL

```
#Generieren Plots von AMI und CoL

#Definieren der Länderdaten
Länderdaten_AMI = [
    {'Länderdaten': df_australia, 'label': 'Australia'},
    {'Länderdaten': df_india, 'label': 'India'},
    {'Länderdaten': df_russia, 'label': 'Russia'},
    {'Länderdaten': df_south_africa, 'label': 'South Africa'},
    {'Länderdaten': df_brazil, 'label': 'Brazil'},
    {'Länderdaten': df_china, 'label': 'China'},
    {'Länderdaten': df_japan, 'label': 'Japan'},
    {'Länderdaten': df_canada, 'label': 'Canada'},
    {'Länderdaten': df_germany, 'label': 'Germany'},
    {'Länderdaten': df_mexico, 'label': 'Mexico'},
    {'Länderdaten': df_united_states, 'label': 'United States'},
    {'Länderdaten': df_france, 'label': 'France'},
]

#Schleifer für die Länderdaten
for länder in Länderdaten_AMI:
    df = länder['Länderdaten']
    label = länder['label']

#Gestalten des Plots
plt.figure()
plt.plot(df["Year"], df['Cost_of_Living'], label='Cost of Living', color='blue')
plt.plot(df["Year"], df['Average_Monthly_Income'], label='Average Monthly Income', color='red')
plt.legend()
plt.xlabel('Year')
plt.ylabel('Amount ' + label)
plt.show()

✓ [10] 737ms
```

ANALYSEANSATZ

- Korrelation AMI & CoL

```
#Korrelation zwischen Cost of Living und Average Monthly Income

#Definieren der Länderdaten
Länderdaten_Korr = [
    {'Länderdaten': df_australia, 'Korrelation': df_australia['Cost_of_Living'].corr(df_australia['Average_Monthly_Income']), 'label': 'Australia' },
    {'Länderdaten': df_india, 'Korrelation': df_india['Cost_of_Living'].corr(df_india['Average_Monthly_Income']), 'label': 'India' },
    {'Länderdaten': df_russia, 'Korrelation': df_russia['Cost_of_Living'].corr(df_russia['Average_Monthly_Income']), 'label': 'Russia' },
    {'Länderdaten': df_south_africa, 'Korrelation': df_south_africa['Cost_of_Living'].corr(df_south_africa['Average_Monthly_Income']), 'label': 'South Africa' },
    {'Länderdaten': df_brazil, 'Korrelation': df_brazil['Cost_of_Living'].corr(df_brazil['Average_Monthly_Income']), 'label': 'Brazil' },
    {'Länderdaten': df_china, 'Korrelation': df_china['Cost_of_Living'].corr(df_china['Average_Monthly_Income']), 'label': 'China' },
    {'Länderdaten': df_japan, 'Korrelation': df_japan['Cost_of_Living'].corr(df_japan['Average_Monthly_Income']), 'label': 'Japan' },
    {'Länderdaten': df_canada, 'Korrelation': df_canada['Cost_of_Living'].corr(df_canada['Average_Monthly_Income']), 'label': 'Canada' },
    {'Länderdaten': df_germany, 'Korrelation': df_germany['Cost_of_Living'].corr(df_germany['Average_Monthly_Income']), 'label': 'Germany' },
    {'Länderdaten': df_mexico, 'Korrelation': df_mexico['Cost_of_Living'].corr(df_mexico['Average_Monthly_Income']), 'label': 'Mexico' },
    {'Länderdaten': df_united_states, 'Korrelation': df_united_states['Cost_of_Living'].corr(df_united_states['Average_Monthly_Income']), 'label': 'United States' },
    {'Länderdaten': df_france, 'Korrelation': df_france['Cost_of_Living'].corr(df_france['Average_Monthly_Income']), 'label': 'France' },
]

for länder in Länderdaten_Korr:
    df = länder['Länderdaten']
    korrelation = länder['Korrelation']
    label = länder['label']

    #Erstellen eines scatterplot mit den Daten aus df
    plt.scatter(df['Average_Monthly_Income'], df['Cost_of_Living'], label = korrelation )

    #Berechnen der Regressionsgeraden
    slope, intercept = np.polyfit(df['Cost_of_Living'], df['Average_Monthly_Income'], 1)
    x = np.linspace(df['Cost_of_Living'].min(), df['Average_Monthly_Income'].max(), 100)
    y = slope * x + intercept

    #Gestalten des Plots
    plt.legend()
    plt.plot(x,y)
    plt.xlabel('Cost of Living ' + label)
    plt.ylabel('Average Monthly Income ' + label)
    plt.show()
```

ANALYSEANSATZ

- Regressionsgeraden
- Year, AMI & CoL

```
#Korrelation von Jahr zu Average_Monthly_Income
#Korrelation von Jahr zu Cost_of_Living
#Definieren der Länderdaten
Länderdaten_Korr_AMI_CoL = [
    {'Länderdaten': df_australia, 'Korrelation_AMI': df_australia['Year'].corr(df_australia['Average_Monthly_Income']), 'Korrelation_CoL': df_australia['Year'].corr(df_australia['Cost_of_Living']), 'label': 'Australia' },
    {'Länderdaten': df_india, 'Korrelation_AMI': df_india['Year'].corr(df_india['Average_Monthly_Income']), 'Korrelation_CoL': df_india['Year'].corr(df_india['Cost_of_Living']), 'label': 'India' },
    {'Länderdaten': df_russia, 'Korrelation_AMI': df_russia['Year'].corr(df_russia['Average_Monthly_Income']), 'Korrelation_CoL': df_russia['Year'].corr(df_russia['Cost_of_Living']), 'label': 'Russia' },
    {'Länderdaten': df_south_africa, 'Korrelation_AMI': df_south_africa['Year'].corr(df_south_africa['Average_Monthly_Income']), 'Korrelation_CoL': df_south_africa['Year'].corr(df_south_africa['Cost_of_Living']), 'label': 'South Africa' },
    {'Länderdaten': df_brazil, 'Korrelation_AMI': df_brazil['Year'].corr(df_brazil['Average_Monthly_Income']), 'Korrelation_CoL': df_brazil['Year'].corr(df_brazil['Cost_of_Living']), 'label': 'Brazil' },
    {'Länderdaten': df_china, 'Korrelation_AMI': df_china['Year'].corr(df_china['Average_Monthly_Income']), 'Korrelation_CoL': df_china['Year'].corr(df_china['Cost_of_Living']), 'label': 'China' },
    {'Länderdaten': df_japan, 'Korrelation_AMI': df_japan['Year'].corr(df_japan['Average_Monthly_Income']), 'Korrelation_CoL': df_japan['Year'].corr(df_japan['Cost_of_Living']), 'label': 'Japan' },
    {'Länderdaten': df_canada, 'Korrelation_AMI': df_canada['Year'].corr(df_canada['Average_Monthly_Income']), 'Korrelation_CoL': df_canada['Year'].corr(df_canada['Cost_of_Living']), 'label': 'Canada' },
    {'Länderdaten': df_germany, 'Korrelation_AMI': df_germany['Year'].corr(df_germany['Average_Monthly_Income']), 'Korrelation_CoL': df_germany['Year'].corr(df_germany['Cost_of_Living']), 'label': 'Germany' },
    {'Länderdaten': df_mexico, 'Korrelation_AMI': df_mexico['Year'].corr(df_mexico['Average_Monthly_Income']), 'Korrelation_CoL': df_mexico['Year'].corr(df_mexico['Cost_of_Living']), 'label': 'Mexico' },
    {'Länderdaten': df_united_states, 'Korrelation_AMI': df_united_states['Year'].corr(df_united_states['Average_Monthly_Income']), 'Korrelation_CoL': df_united_states['Year'].corr(df_united_states['Cost_of_Living']), 'label': 'United' },
    {'Länderdaten': df_france, 'Korrelation_AMI': df_france['Year'].corr(df_france['Average_Monthly_Income']), 'Korrelation_CoL': df_france['Year'].corr(df_france['Cost_of_Living']), 'label': 'France' },
]

for länder in Länderdaten_Korr_AMI_CoL:
    df = länder['Länderdaten']
    korrelation_ami = länder['Korrelation_AMI']
    korrelation_col = länder['Korrelation_CoL']
    label = länder['label']

    plt.scatter(df['Year'], df['Average_Monthly_Income'], label='Average Monthly Income', color='blue')
    slope_AMI, intercept_AMI = np.polyfit(df['Year'], df['Average_Monthly_Income'], 1)
    plt.plot(df['Year'], slope_AMI * df['Year'] + intercept_AMI, color='blue')

    plt.scatter(df['Year'], df['Cost_of_Living'], label='Cost of Living', color='red')
    slope_CoL, intercept_CoL = np.polyfit(df['Year'], df['Cost_of_Living'], 1)
    plt.plot(df['Year'], slope_CoL * df['Year'] + intercept_CoL, color='red')

    plt.xlabel('Year')
    plt.ylabel('Amount')
    plt.title(label)
    plt.legend()
    plt.show()
```


VERHÄLTNIS VON EINNAHMEN UND AUSGABE

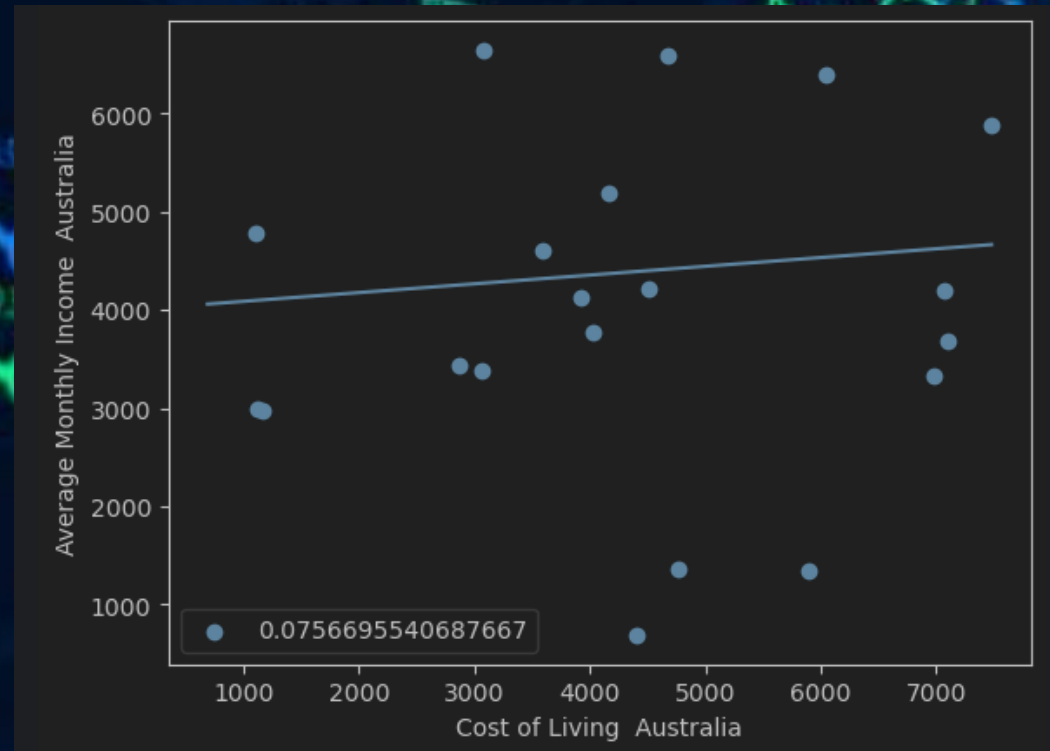
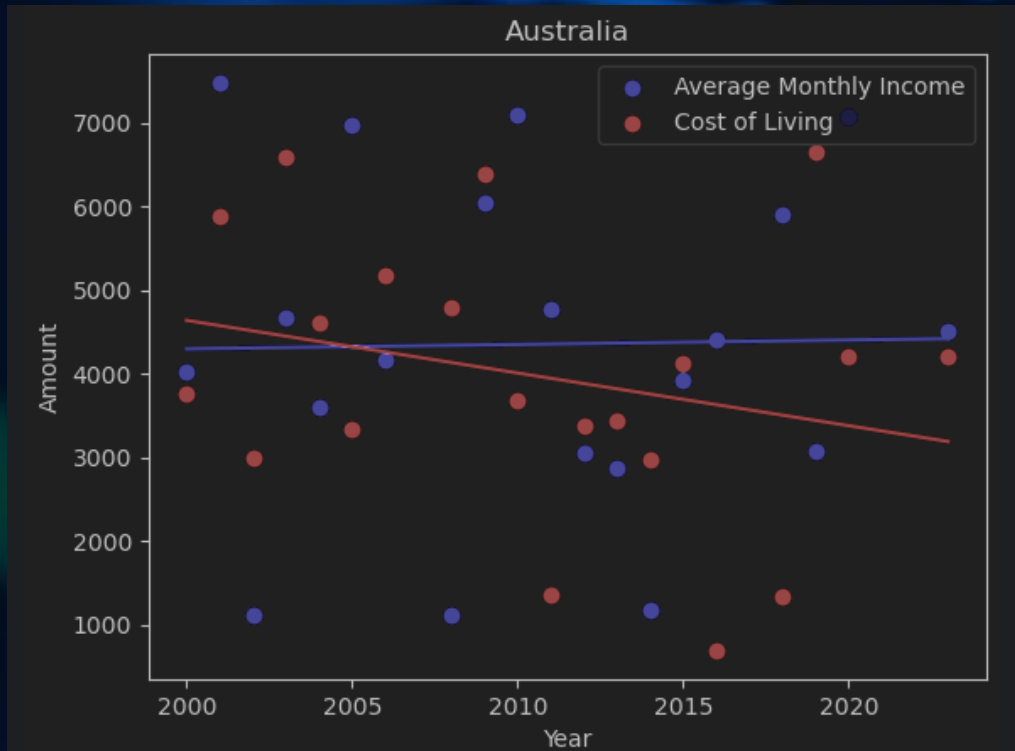
- Australien



REGRESSION

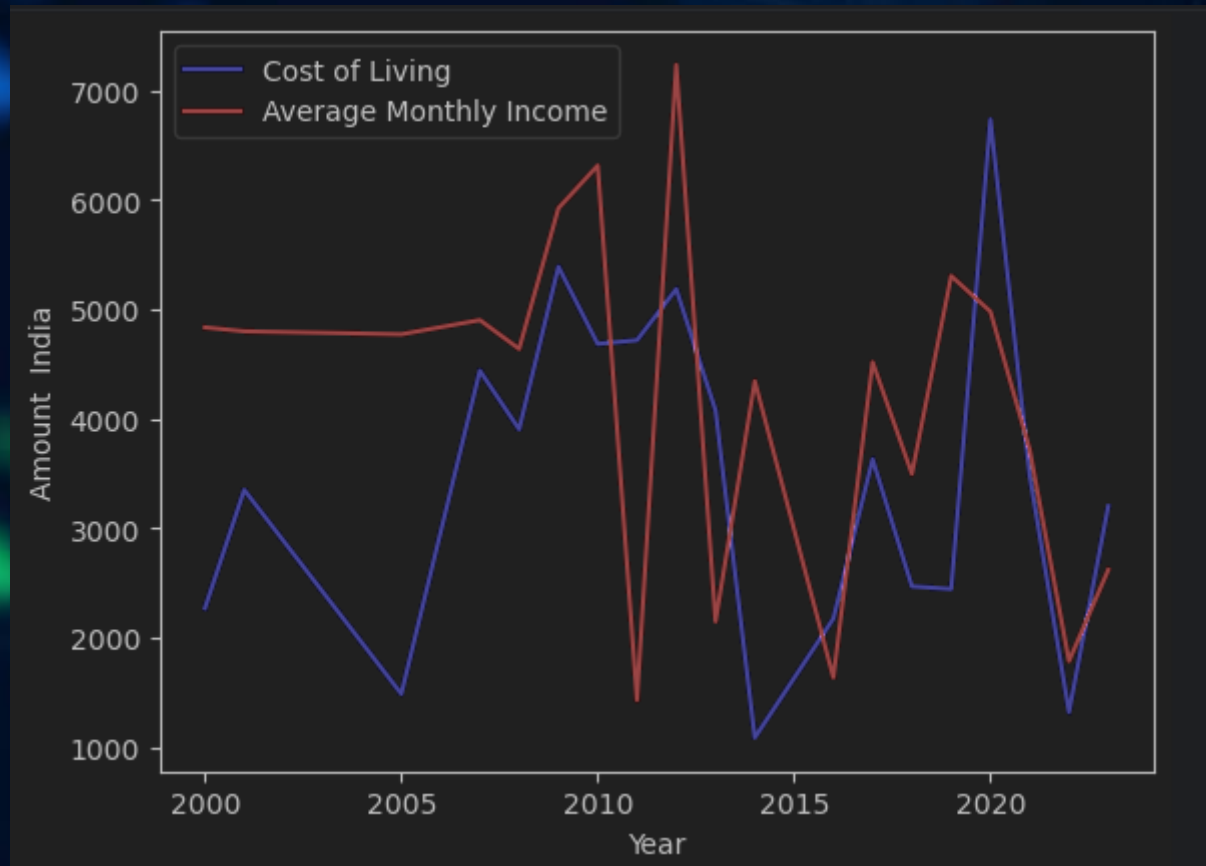
- Australien

0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



VERHÄLTNIS VON EINNAHMEN UND AUSGABE

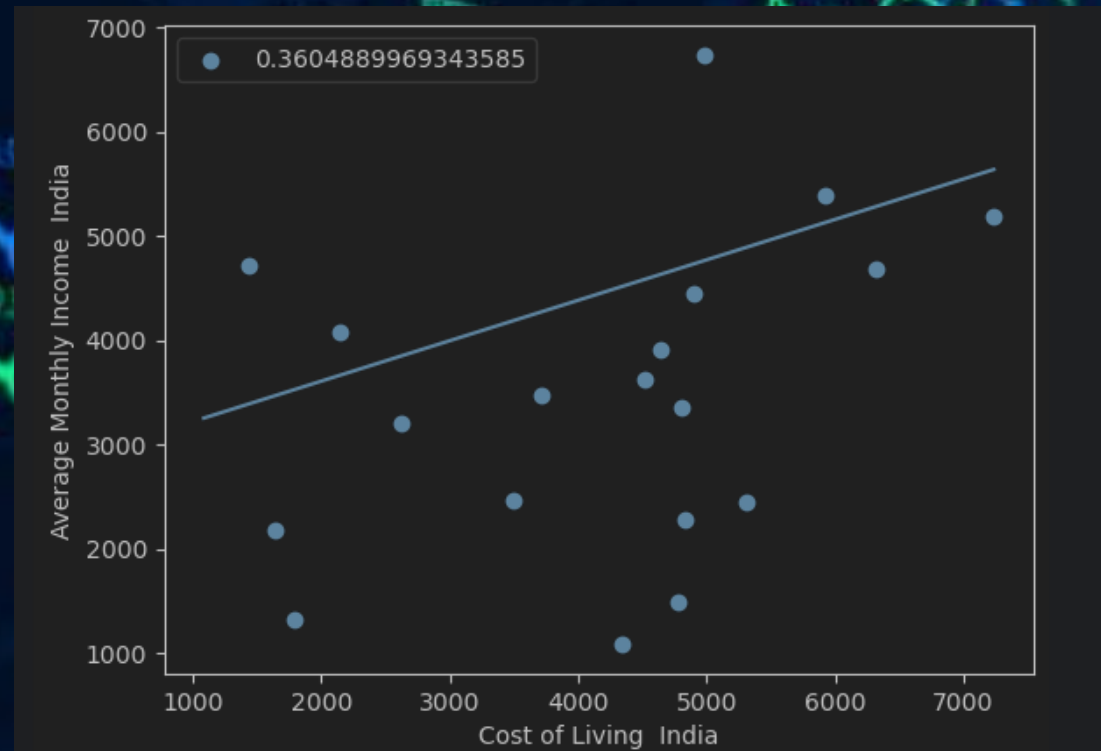
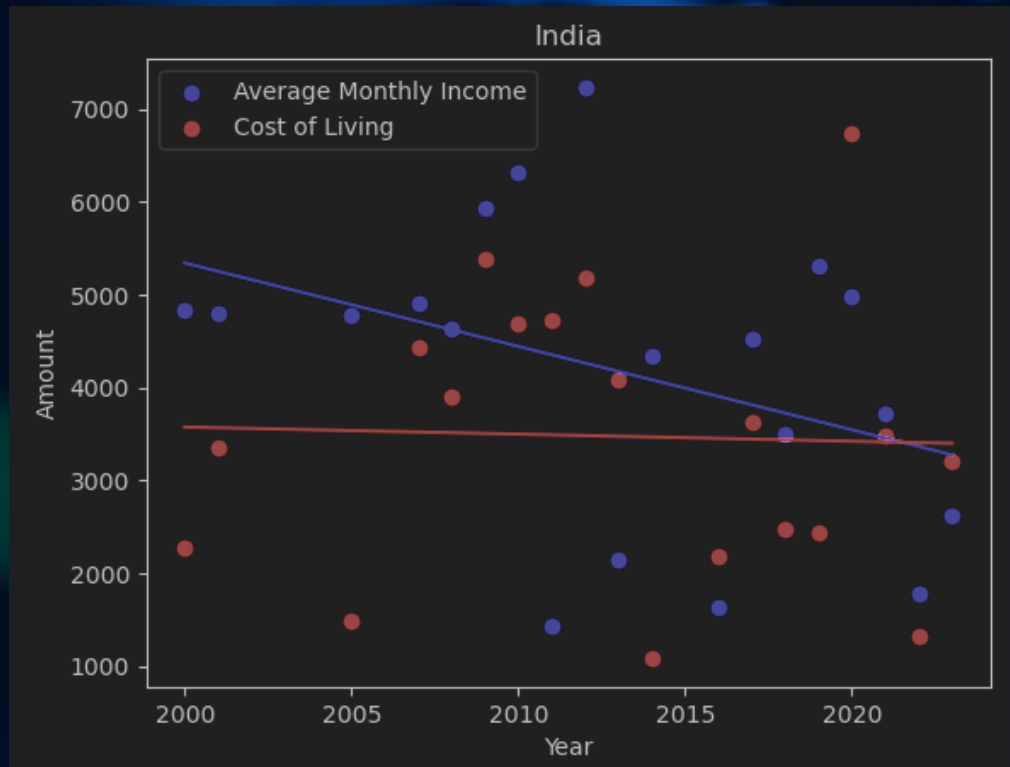
- India



REGRESSION

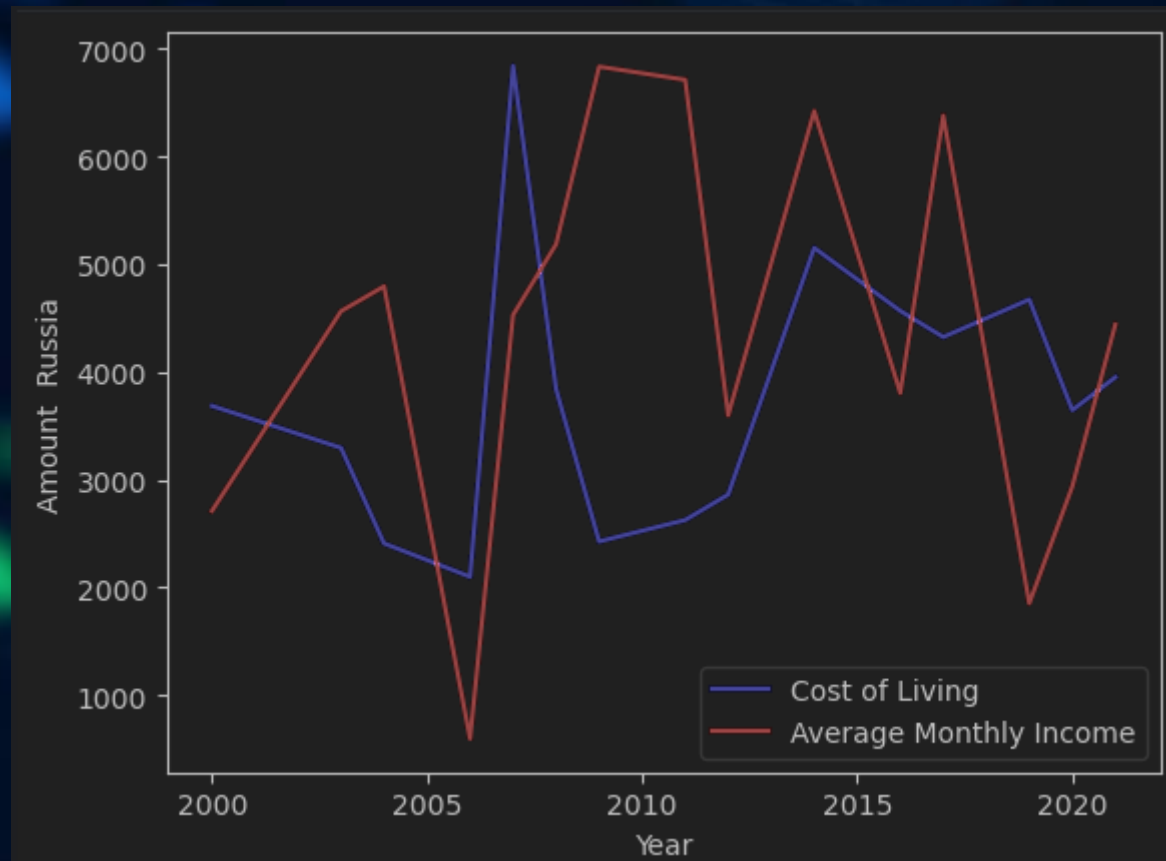
- India

0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



VERHÄLTNIS VON EINNAHMEN UND AUSGABE

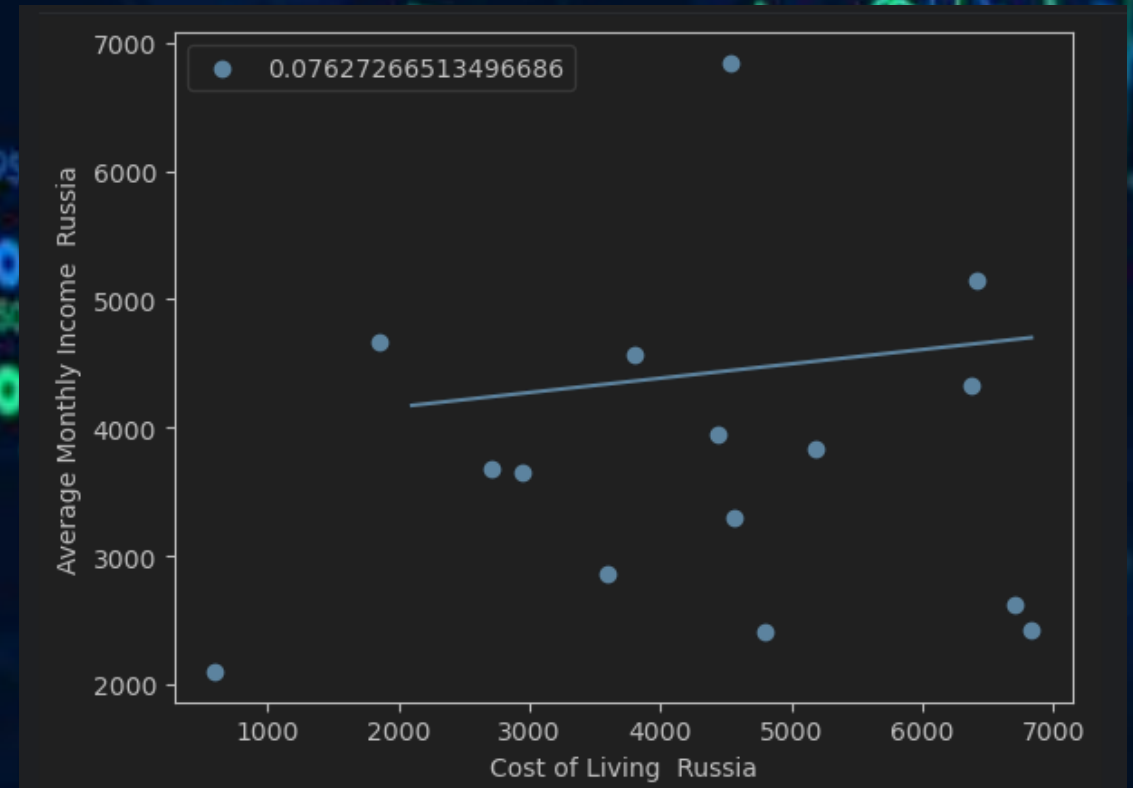
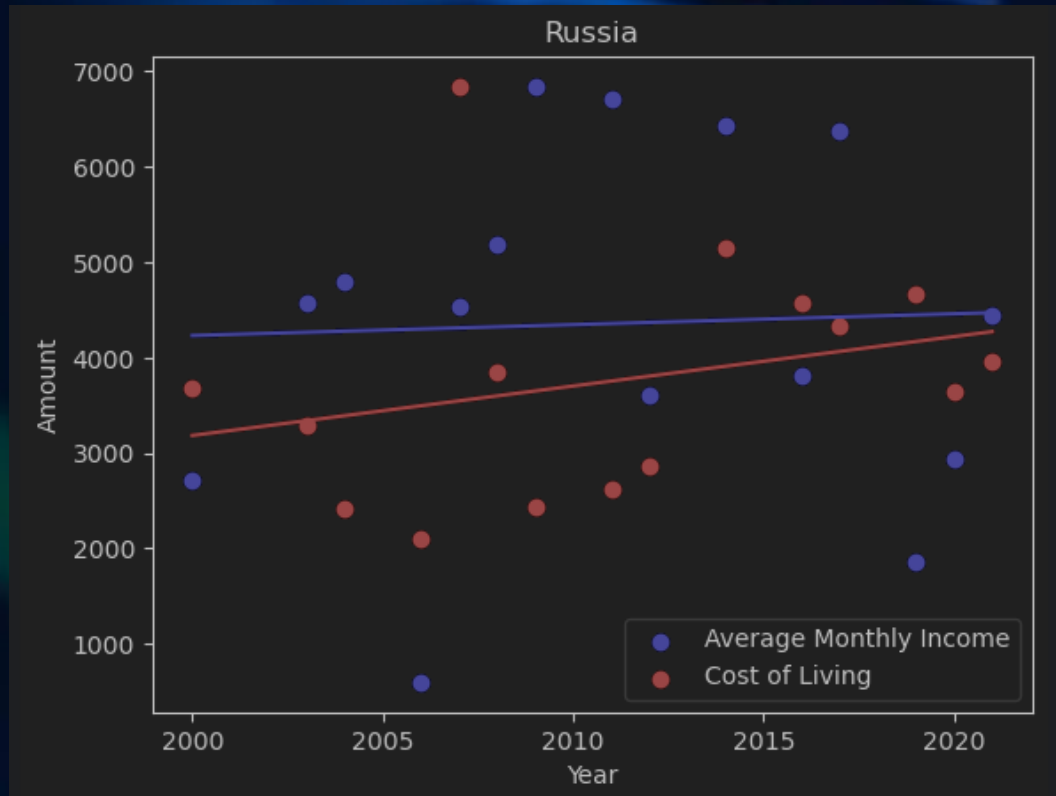
- Russia



REGRESSION

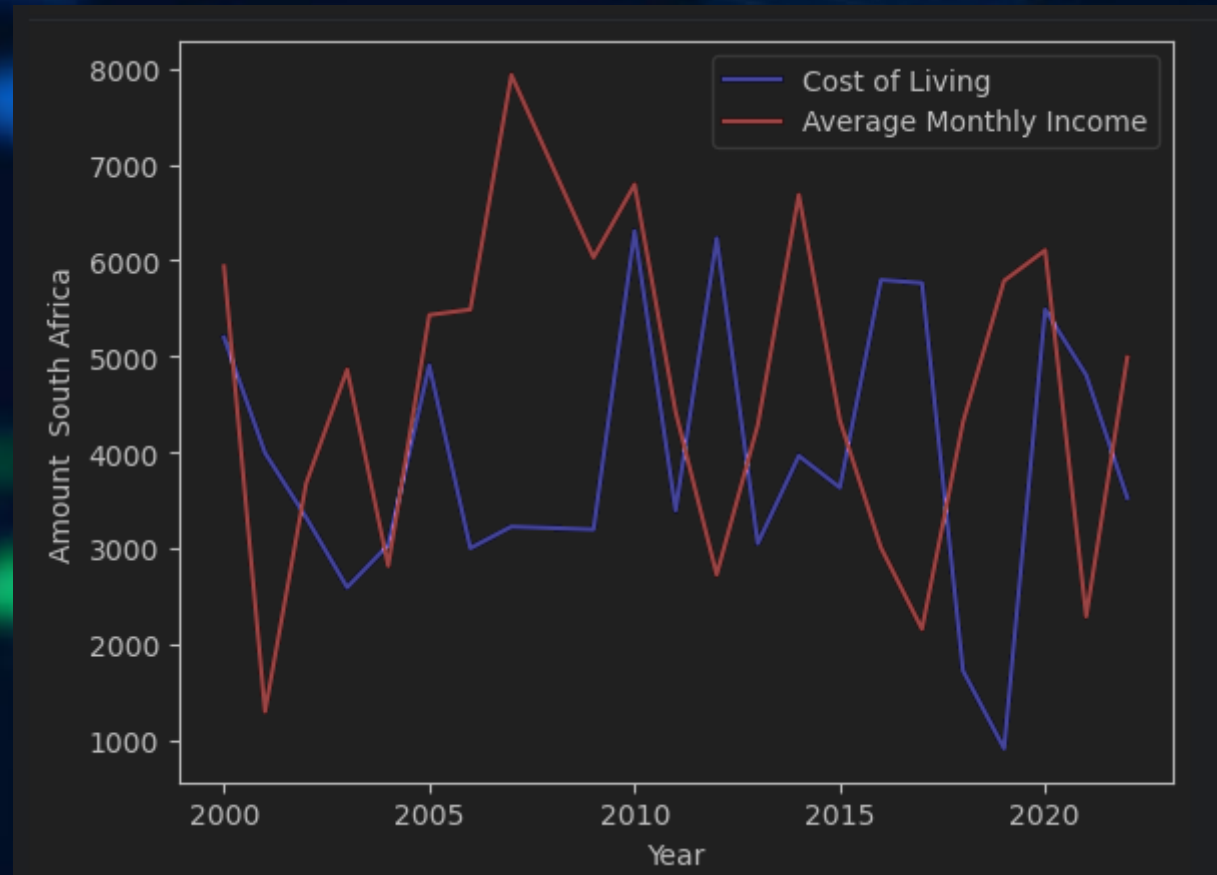
- Russia

0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



VERHÄLTNIS VON EINNAHMEN UND AUSGABE

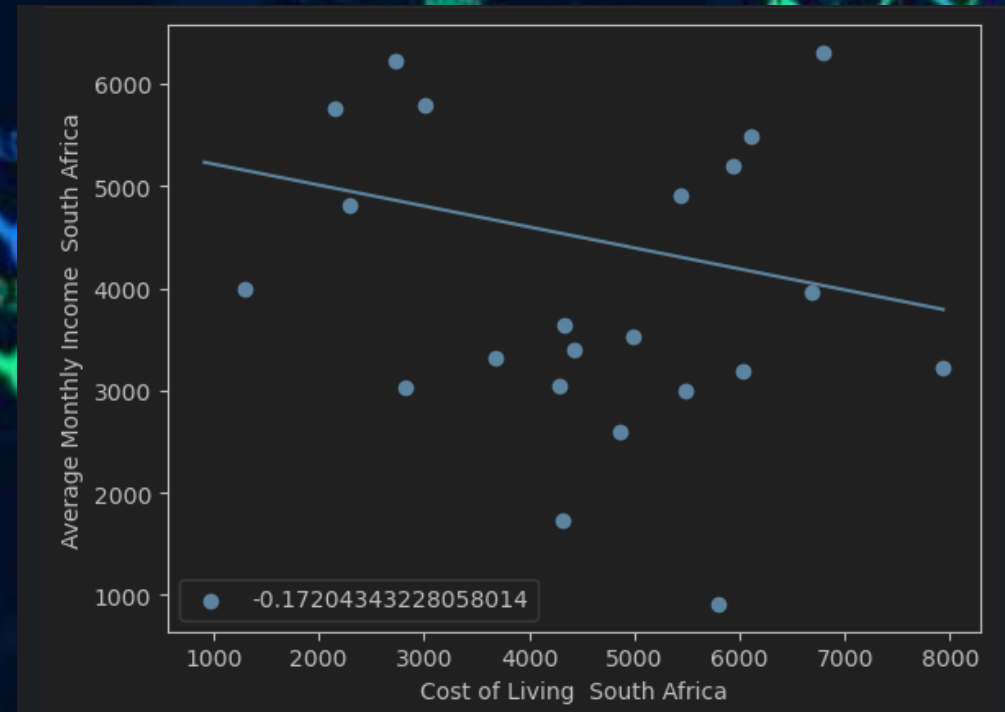
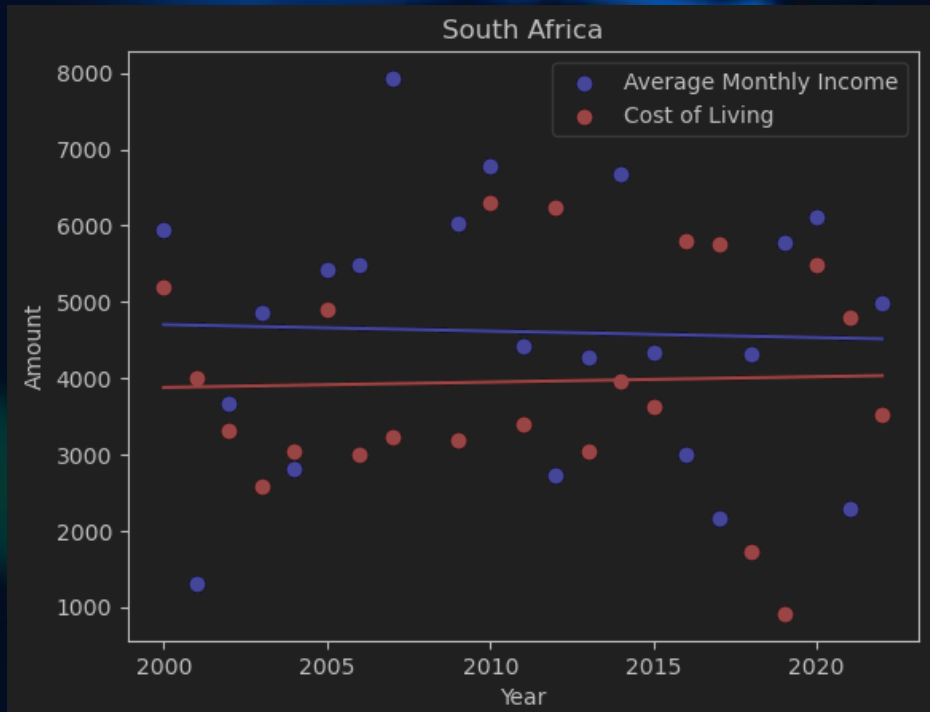
- South Africa



REGRESSION

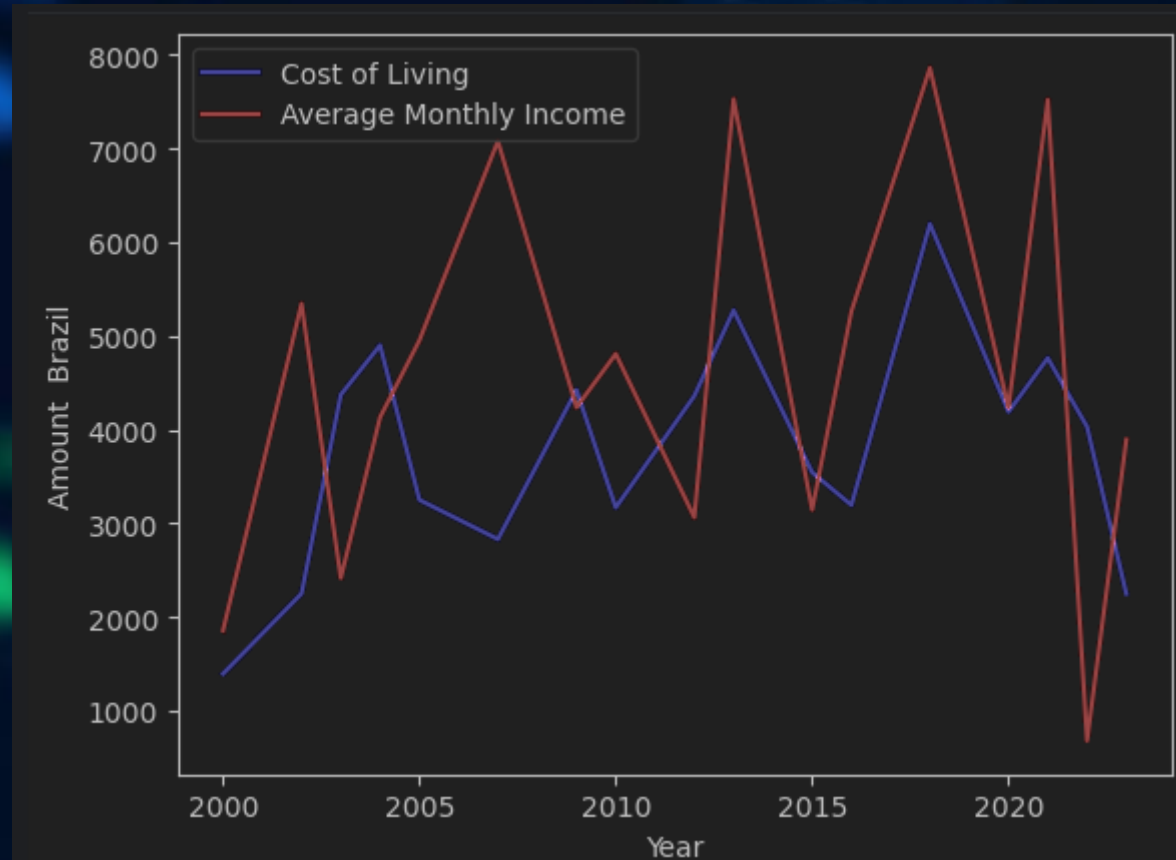
- South Africa

0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



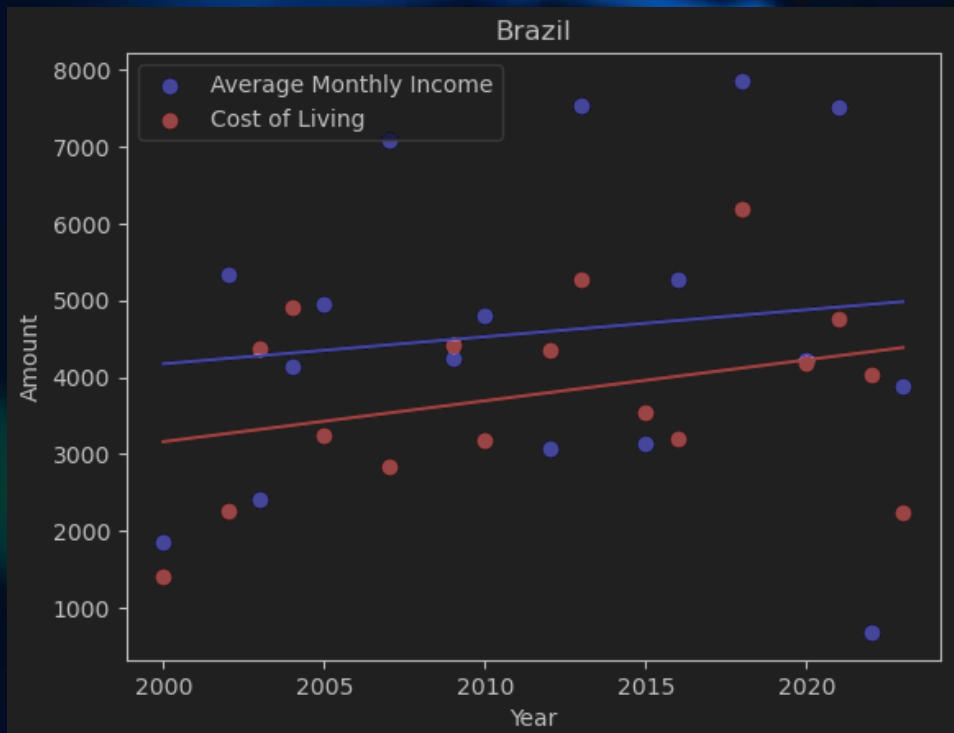
VERHÄLTNIS VON EINNAHMEN UND AUSGABE

- Brazil

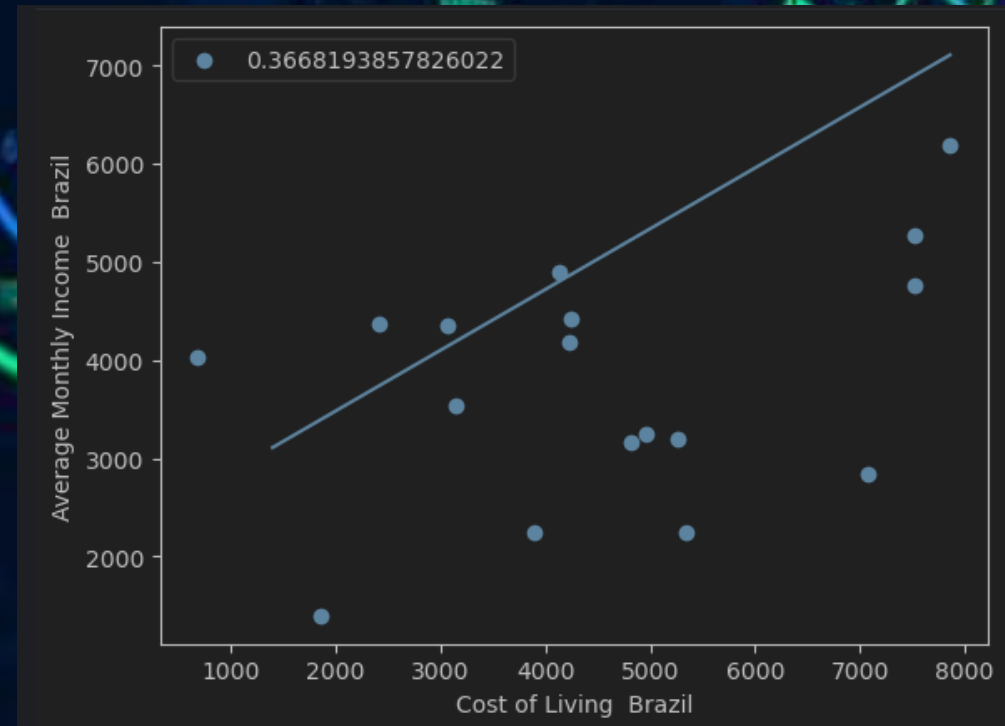


REGRESSION

- Brazil

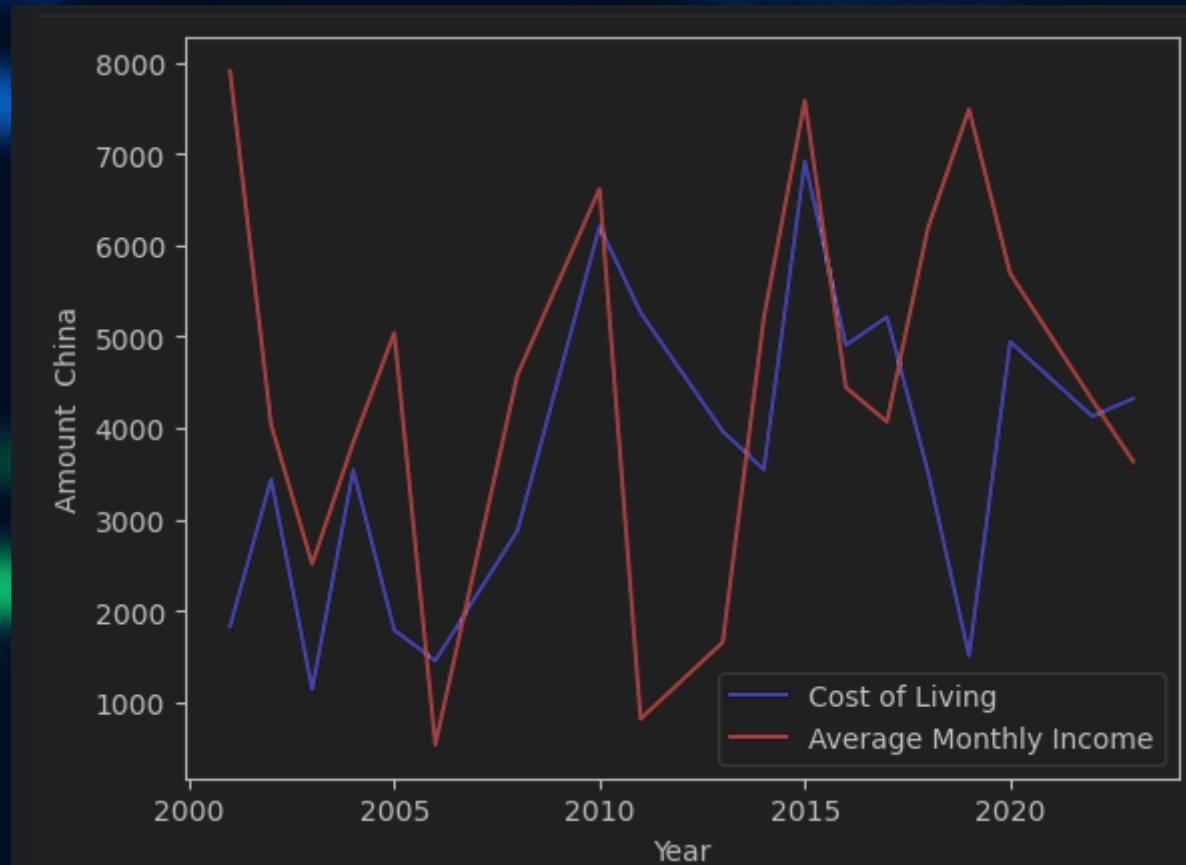


0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



VERHÄLTNIS VON EINNAHMEN UND AUSGABE

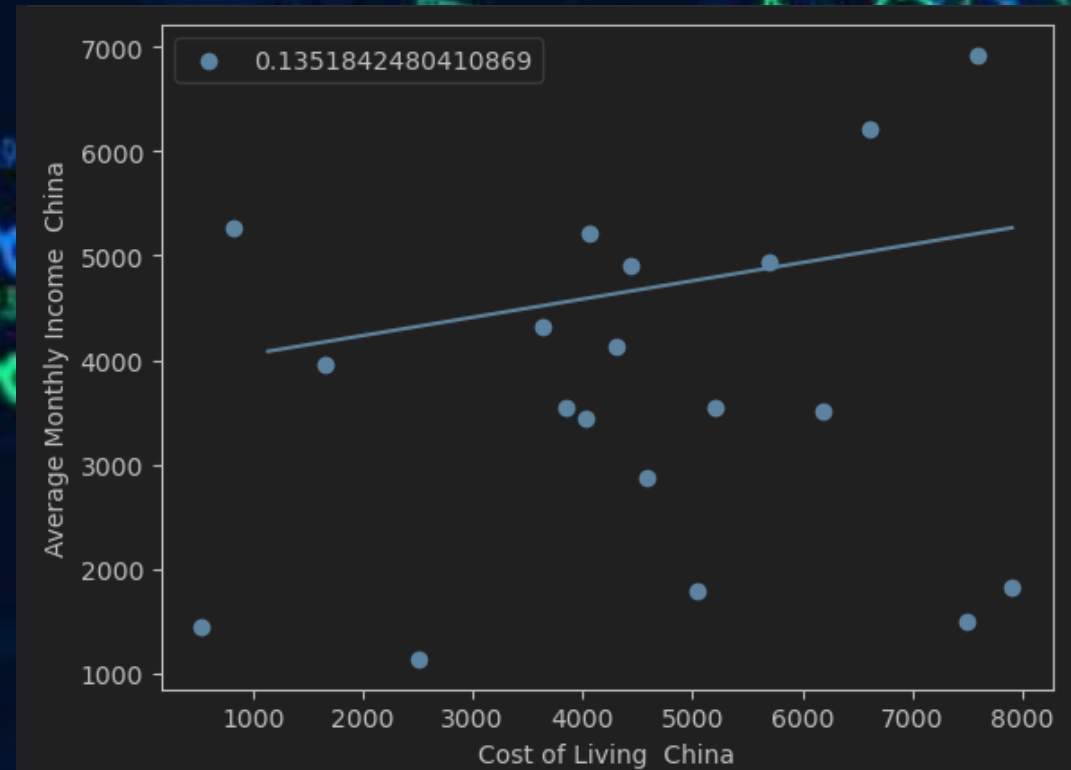
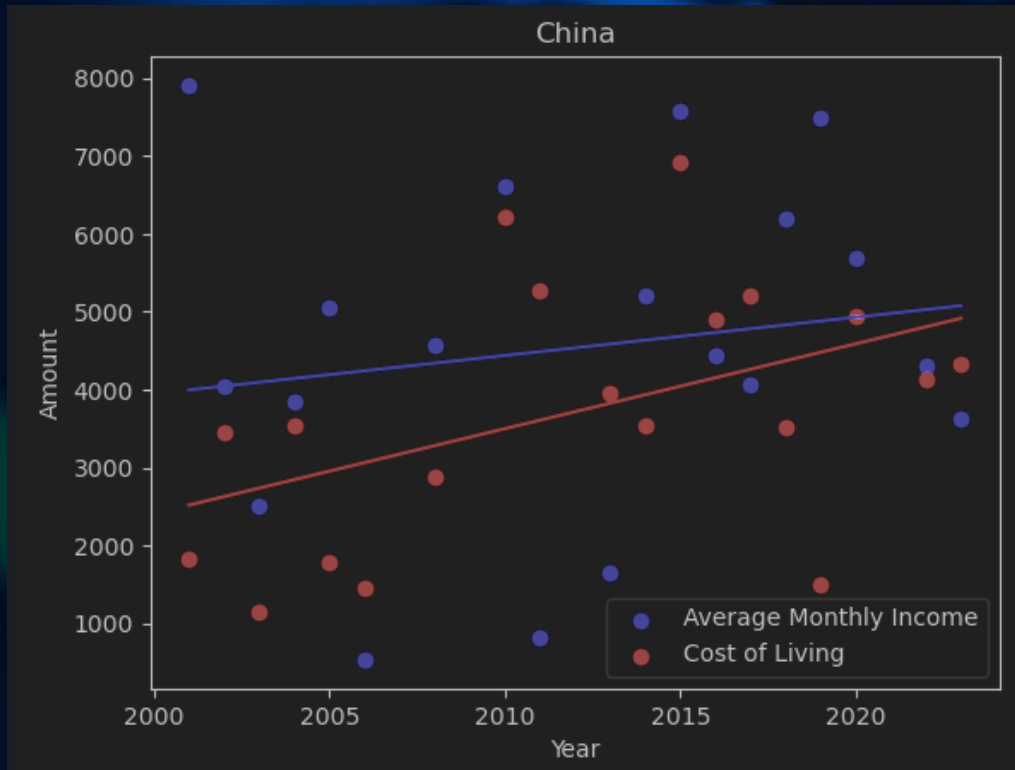
- China



REGRESSION

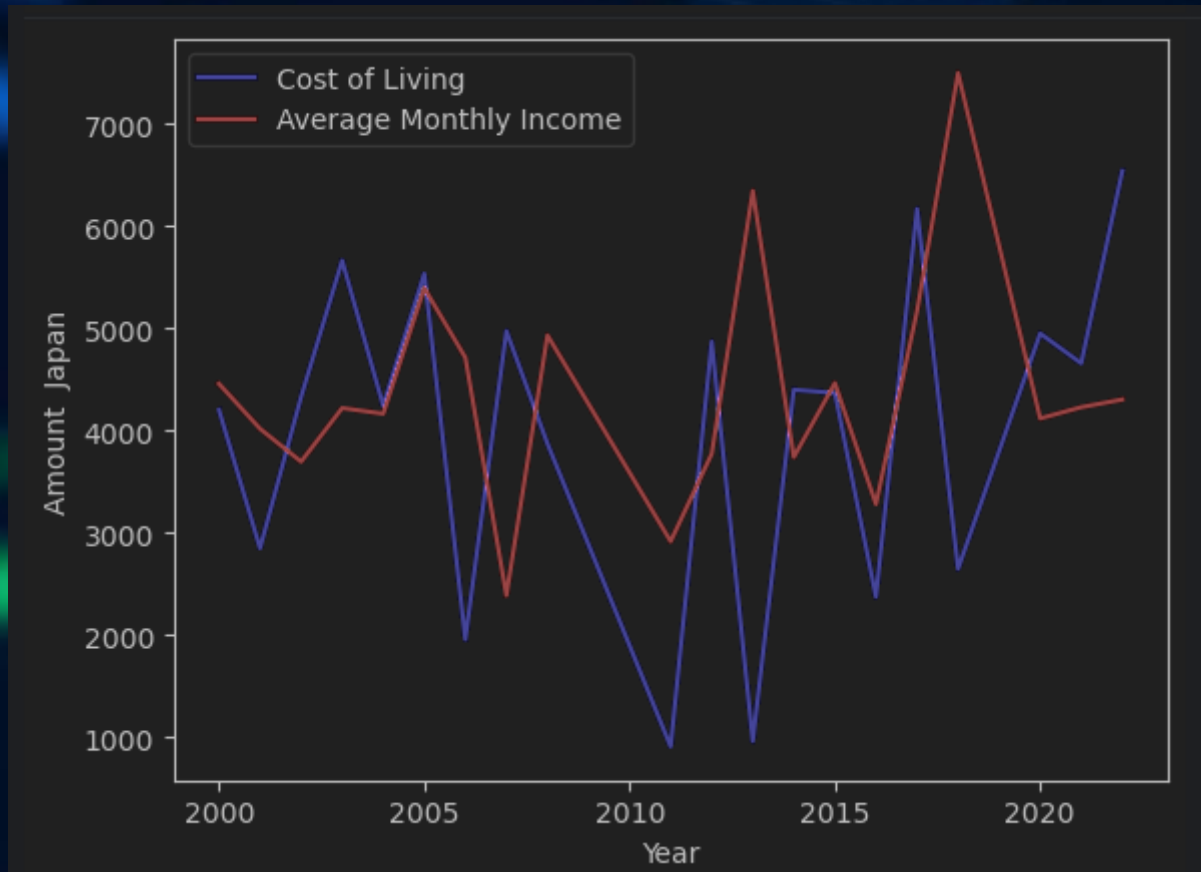
- China

0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



VERHÄLTNIS VON EINNAHMEN UND AUSGABE

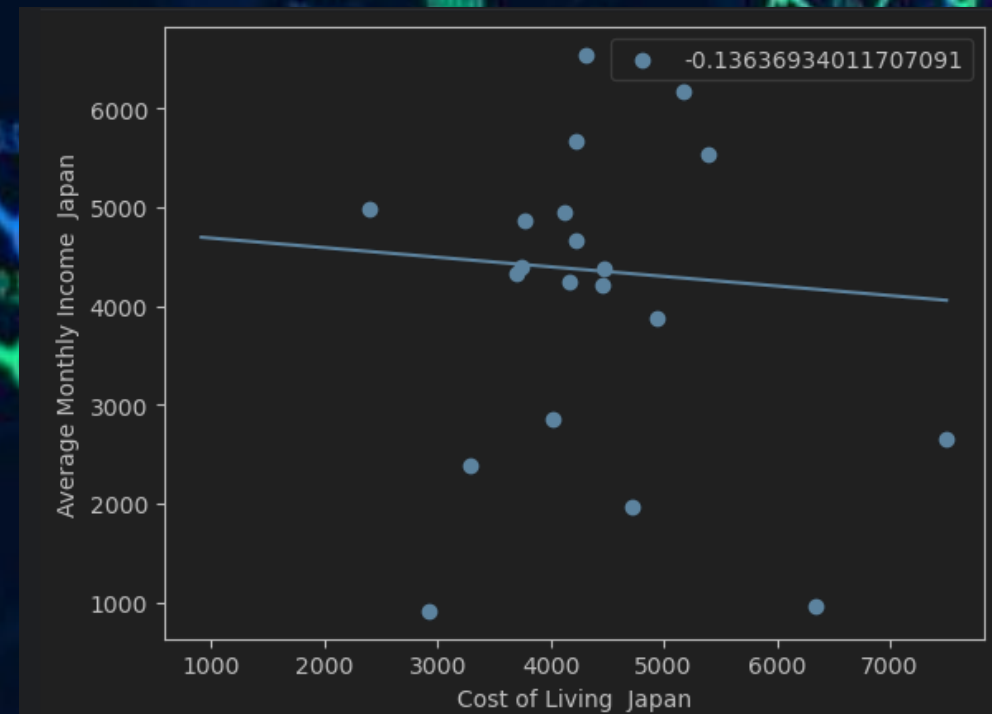
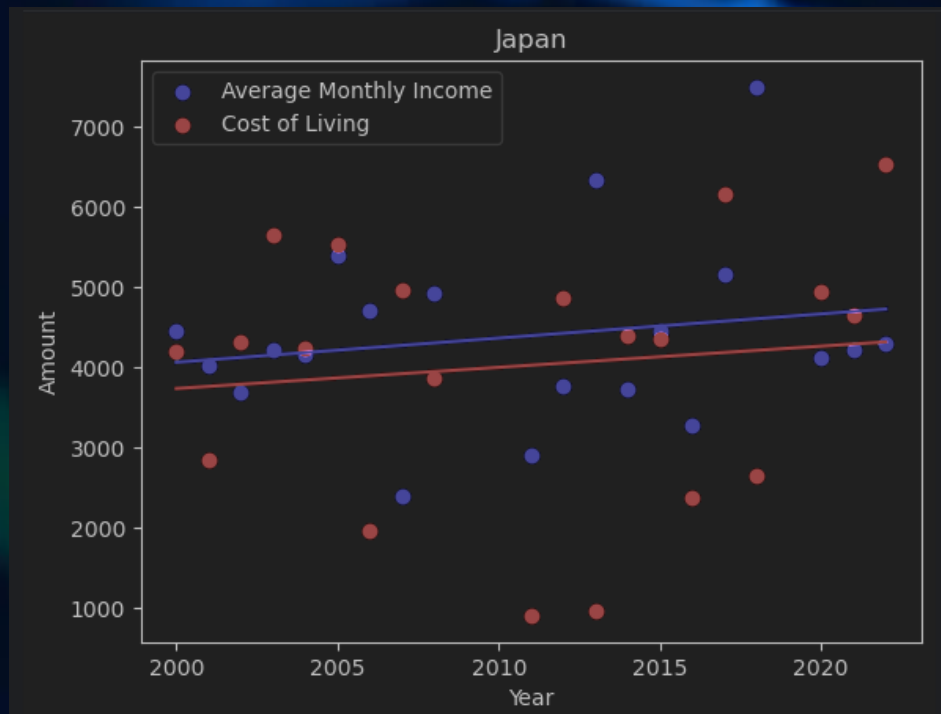
- Japan



REGRESSION

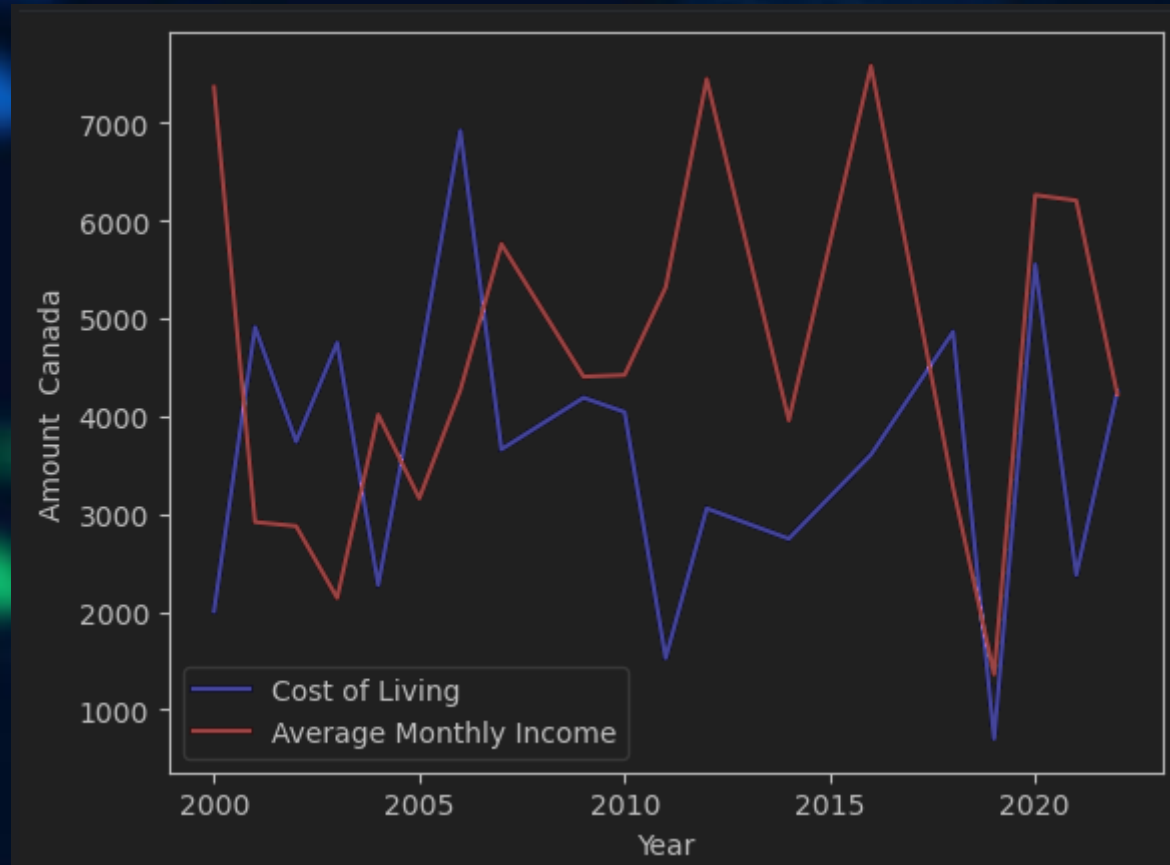
- Japan

0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



VERHÄLTNIS VON EINNAHMEN UND AUSGABE

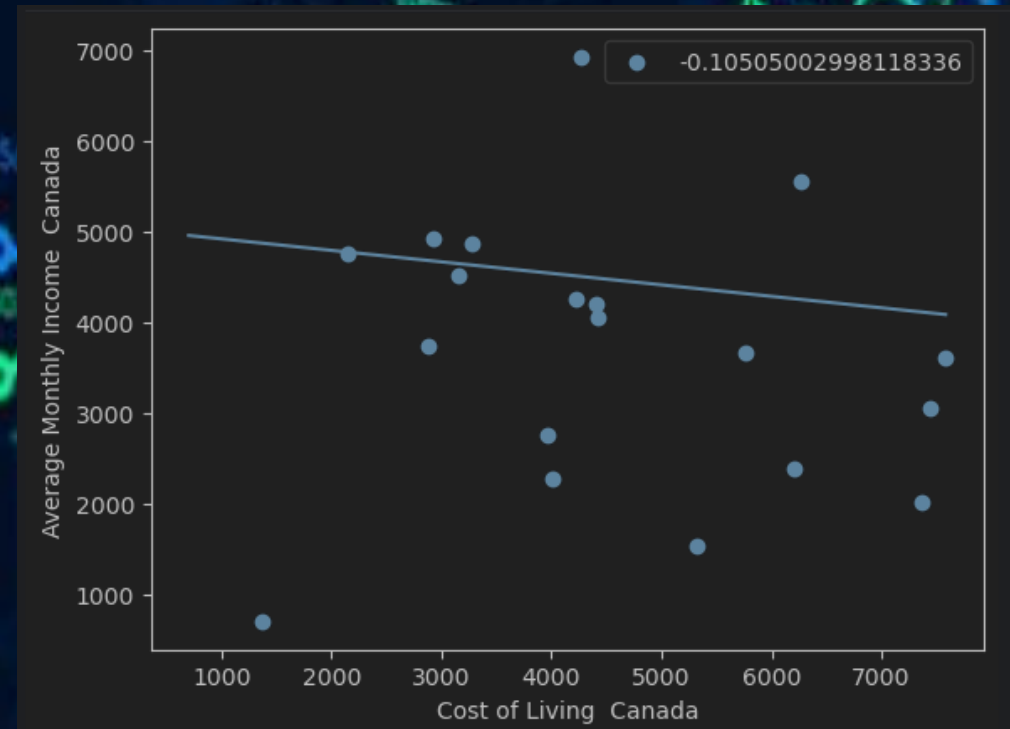
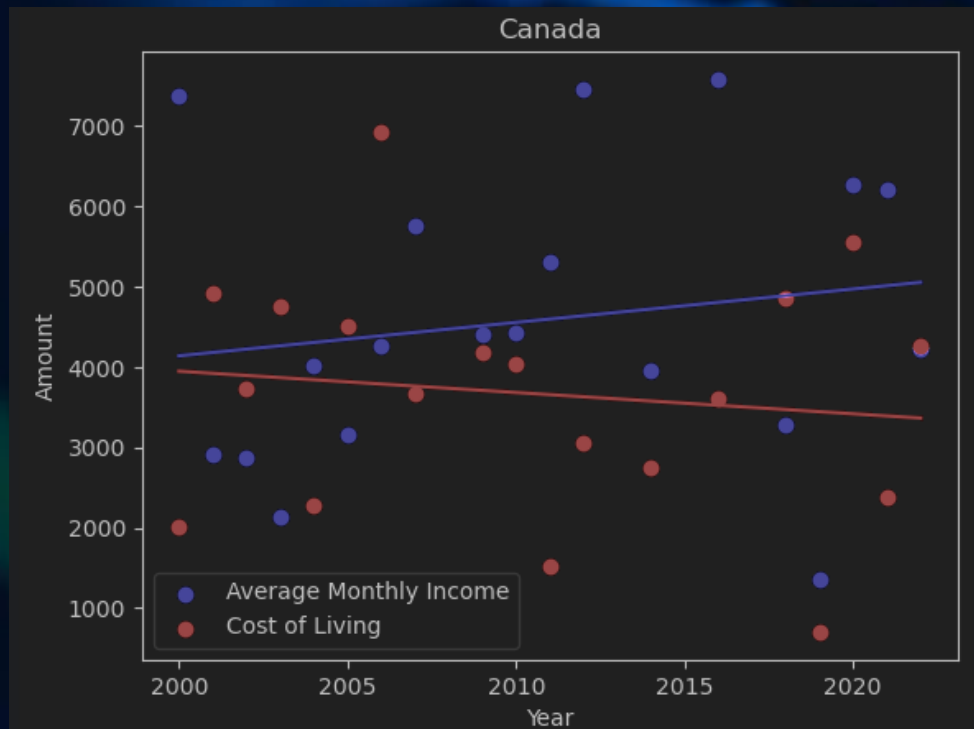
- Canada



REGRESSION

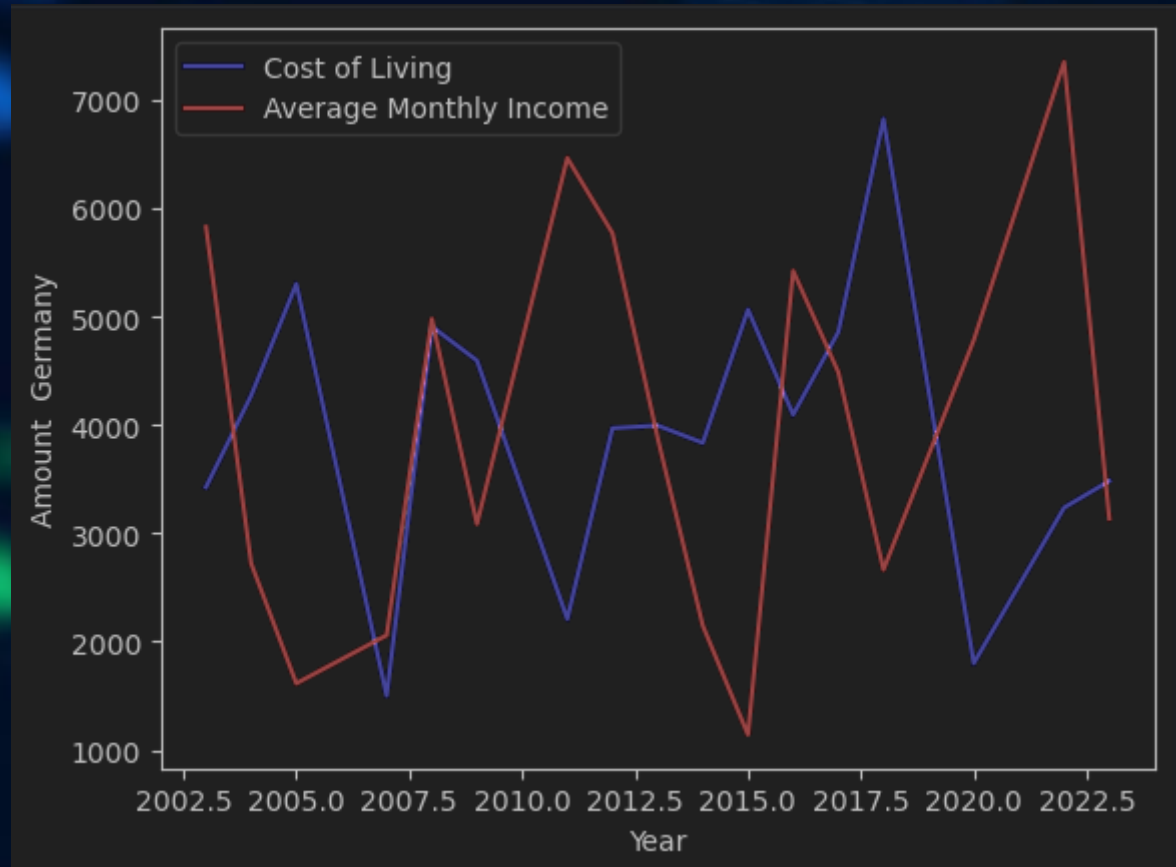
- Canada

0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



VERHÄLTNIS VON EINNAHMEN UND AUSGABE

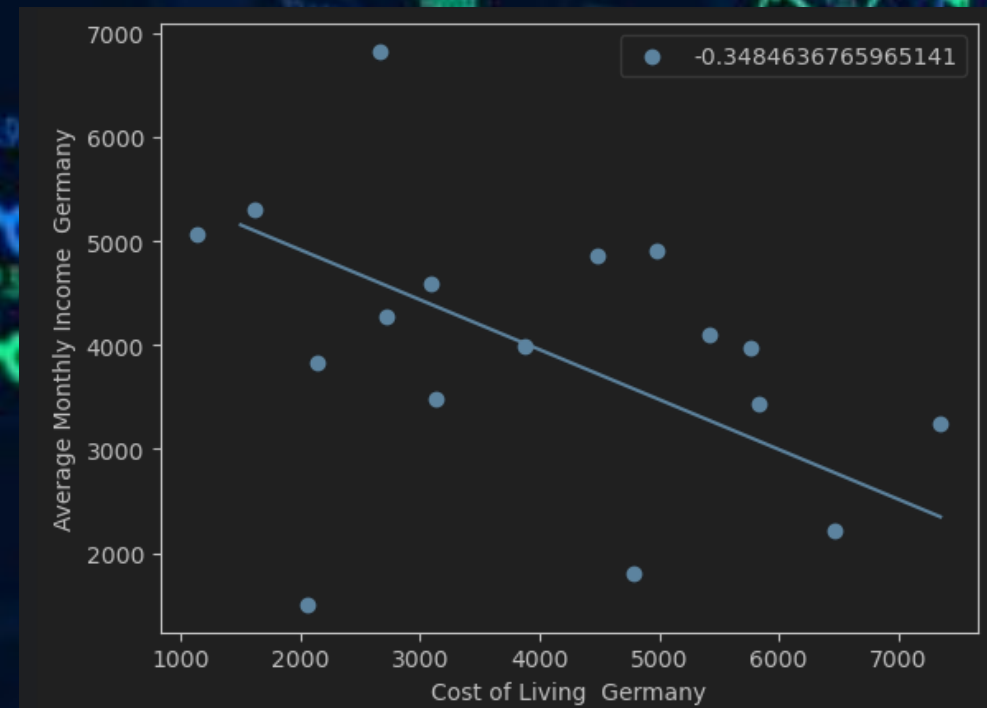
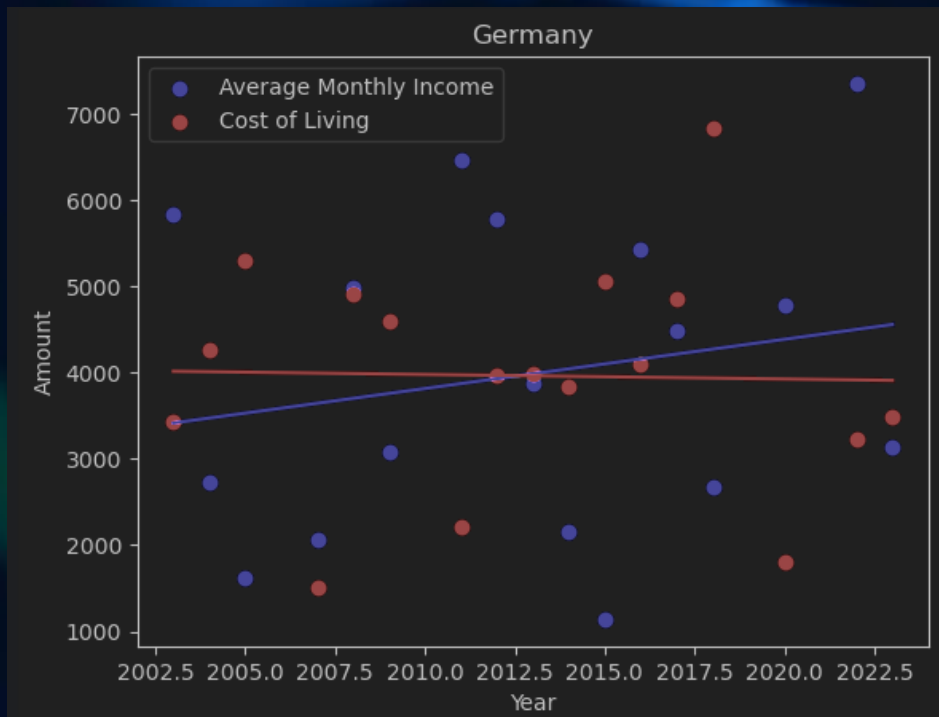
- Germany



REGRESSION

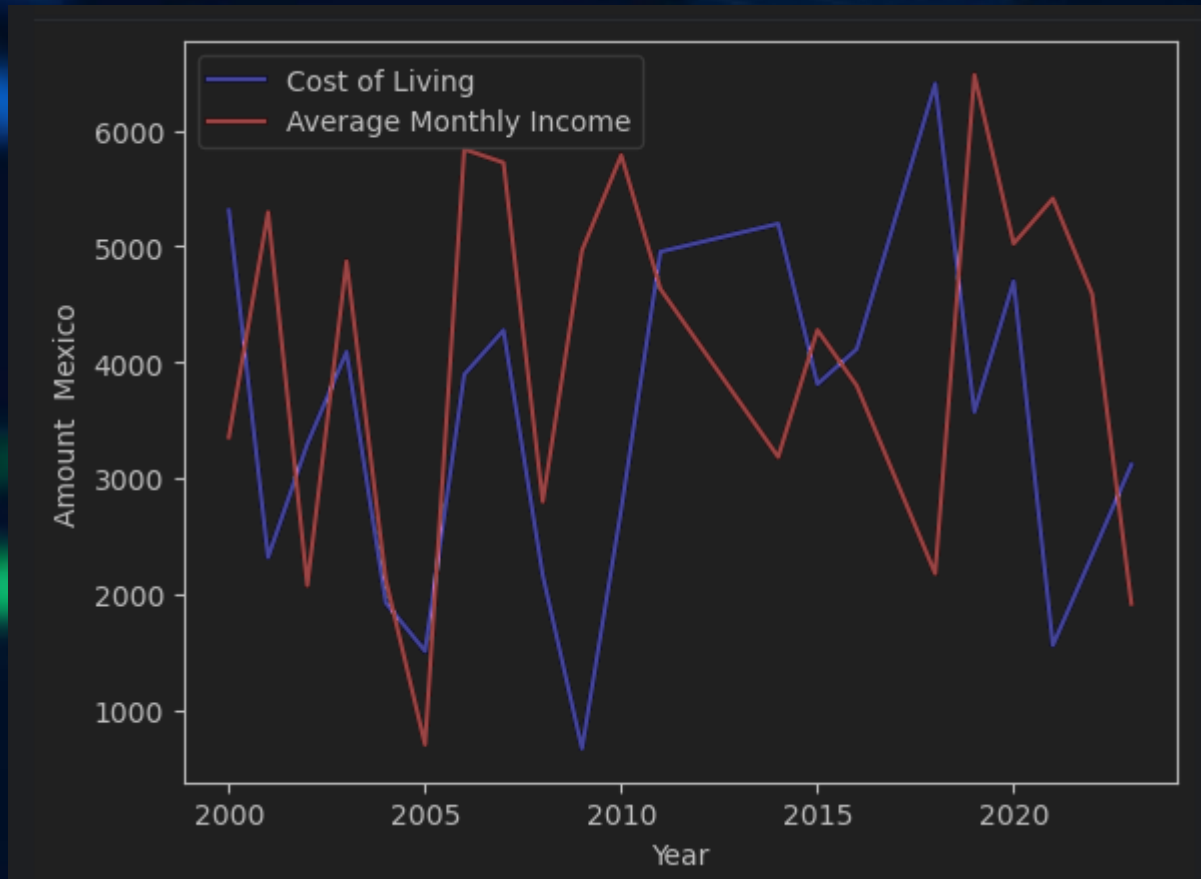
- Germany

0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



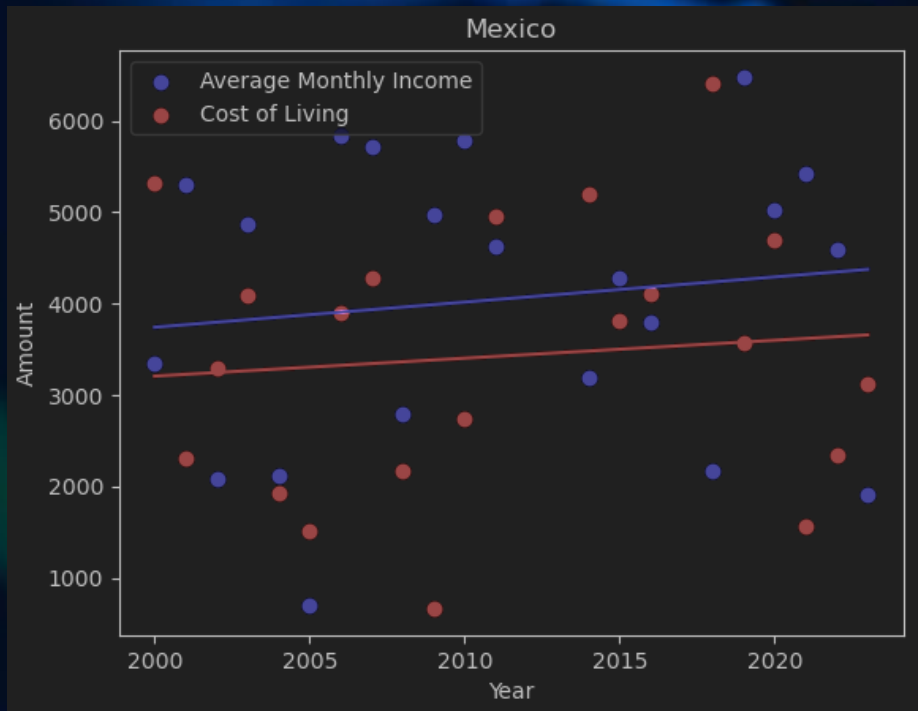
VERHÄLTNIS VON EINNAHMEN UND AUSGABE

- Mexico

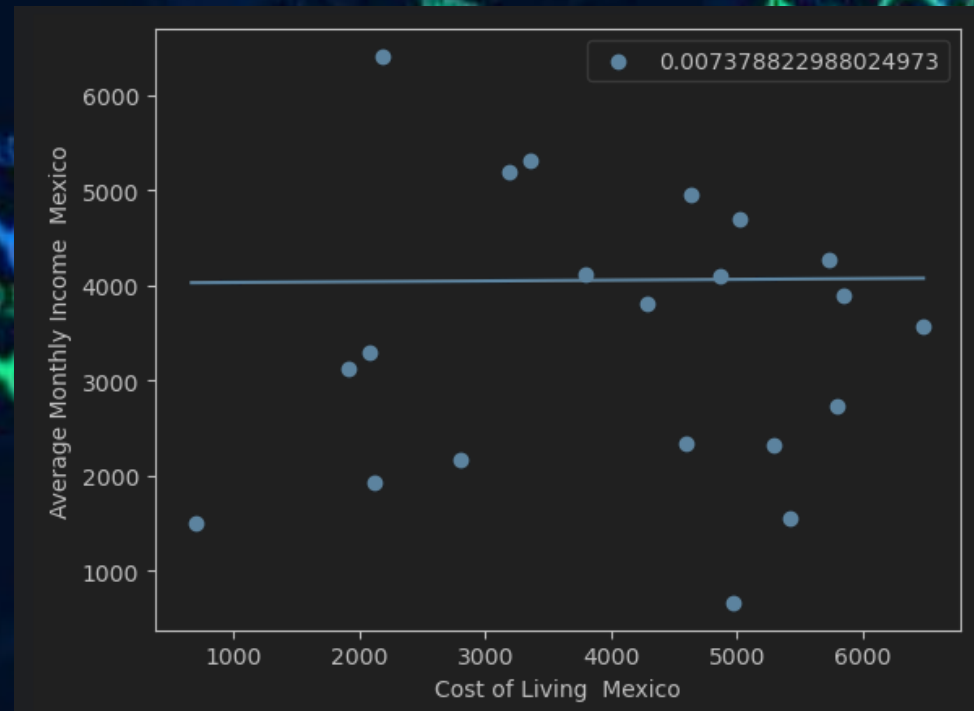


REGRESSION

- Mexico



0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



VERHÄLTNIS VON EINNAHMEN UND AUSGABE

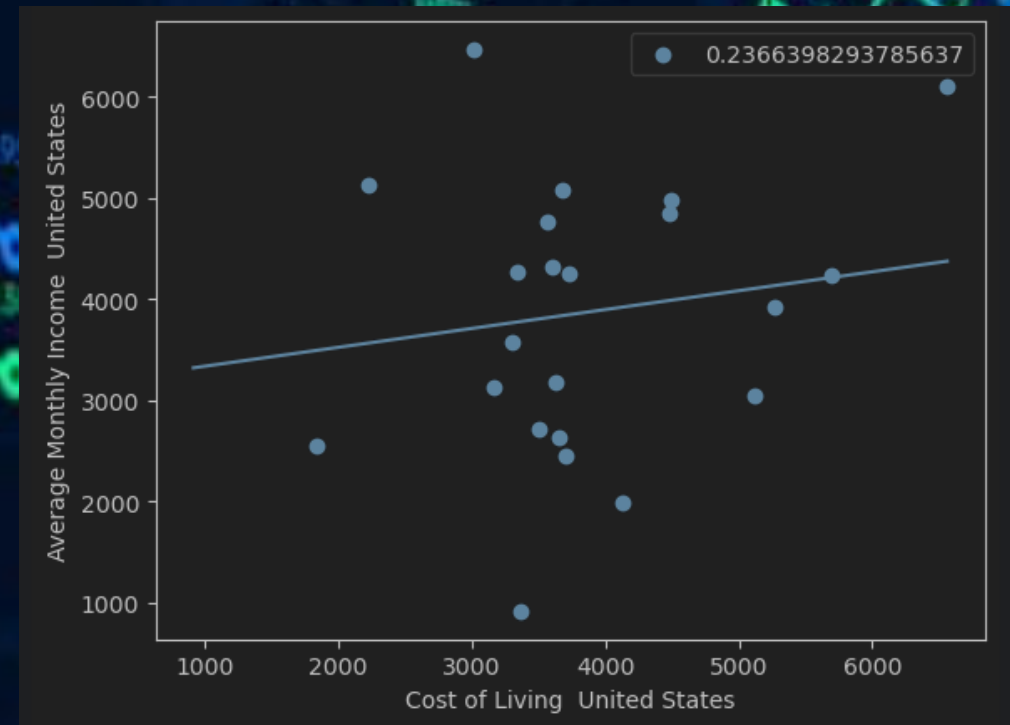
- United States



REGRESSION

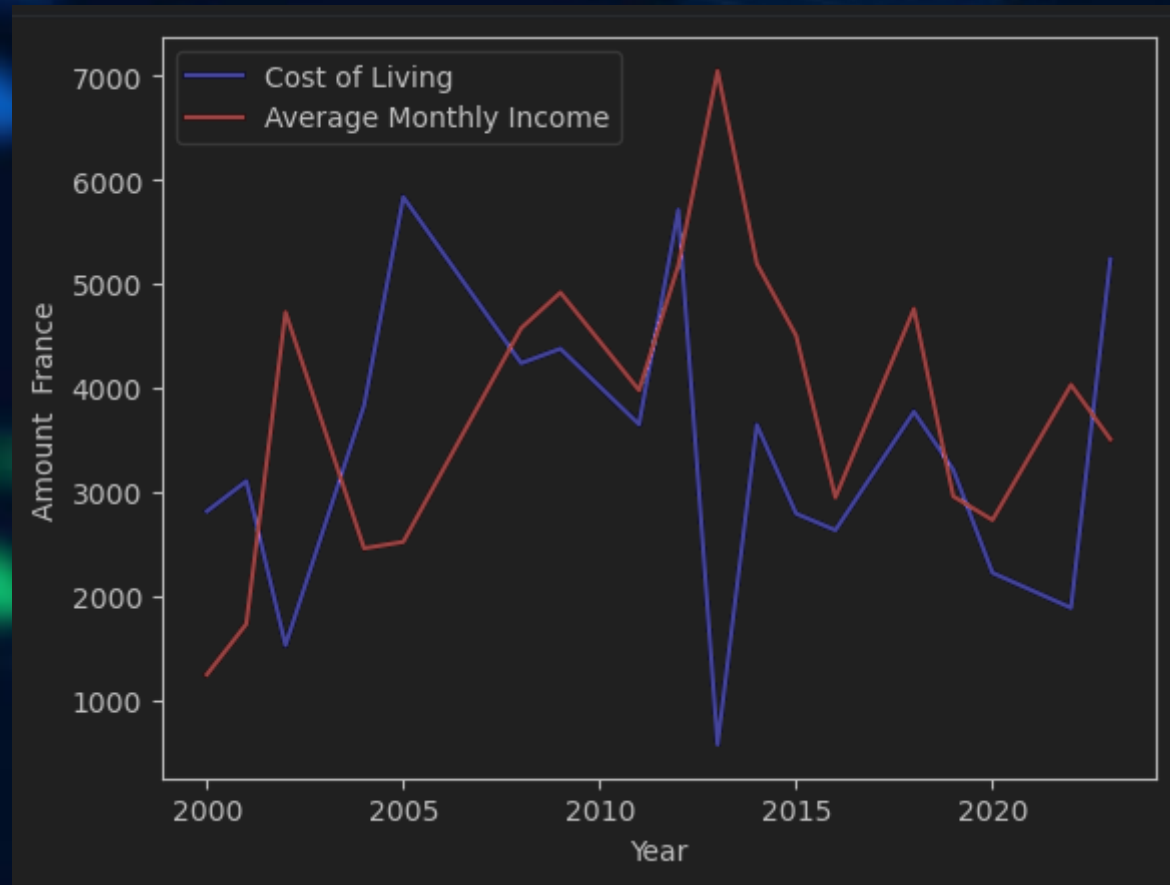
- United States

0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



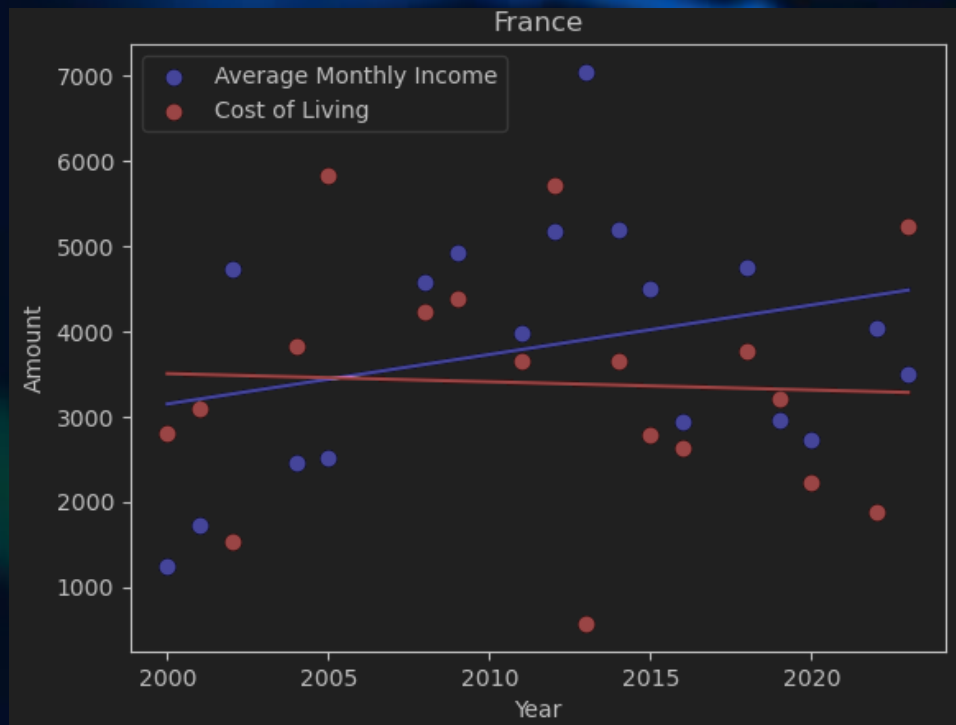
VERHÄLTNIS VON EINNAHMEN UND AUSGABE

- France

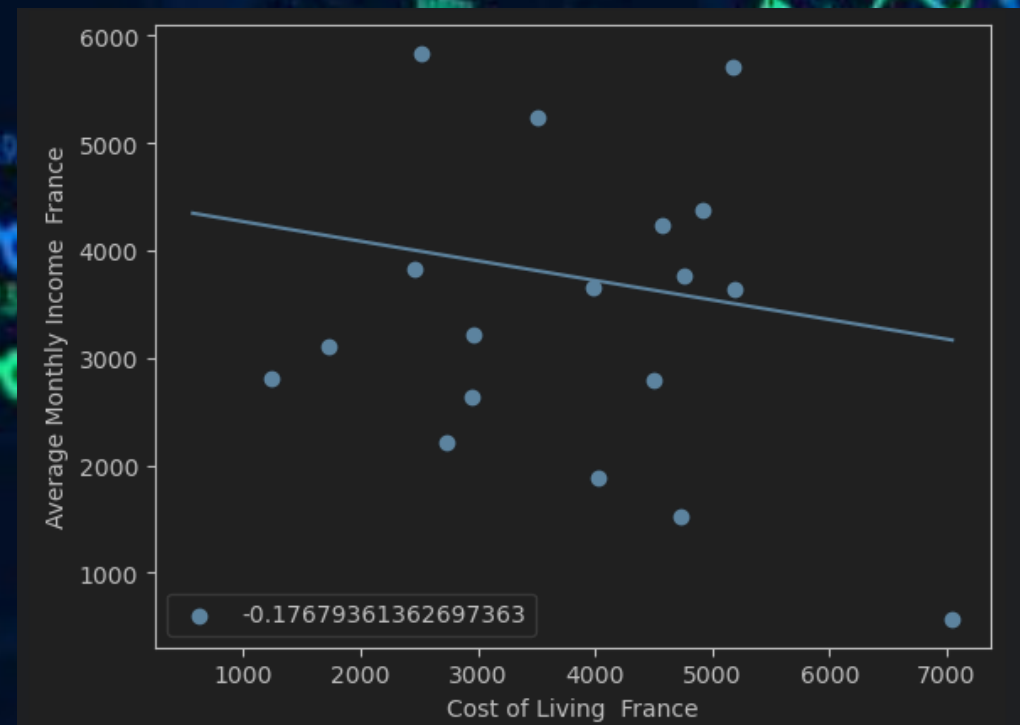


REGRESSION

- France

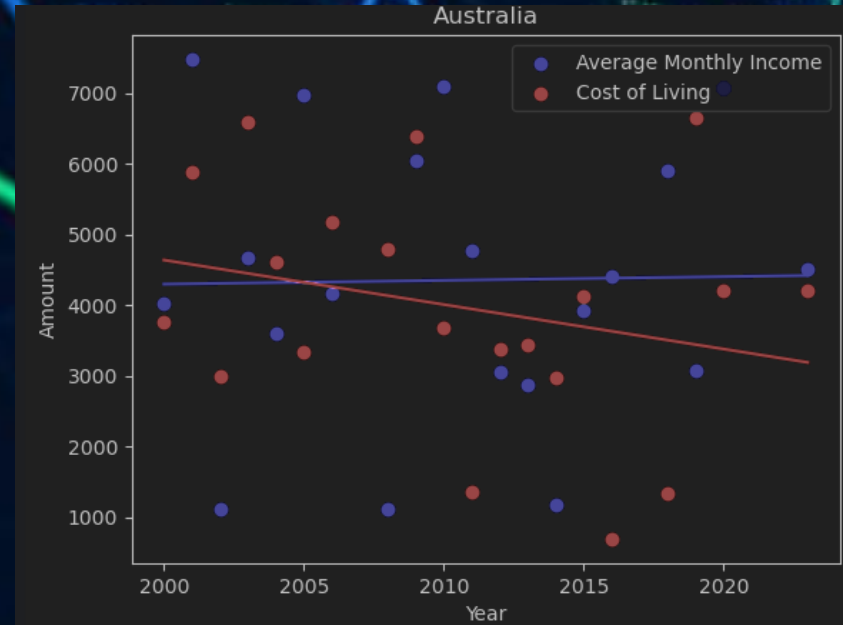
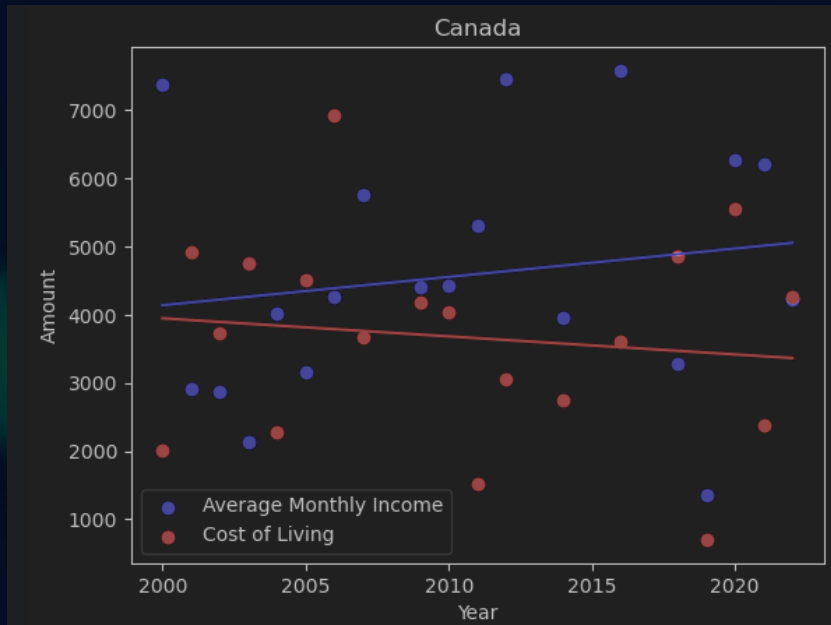


0,0 < 0,1 kein Zusammenhang
0,1 < 0,3 geringer Zusammenhang
0,3 < 0,5 mittlerer Zusammenhang
0,5 < 0,7 hoher Zusammenhang
0,7 < 1 sehr hoher Zusammenhang



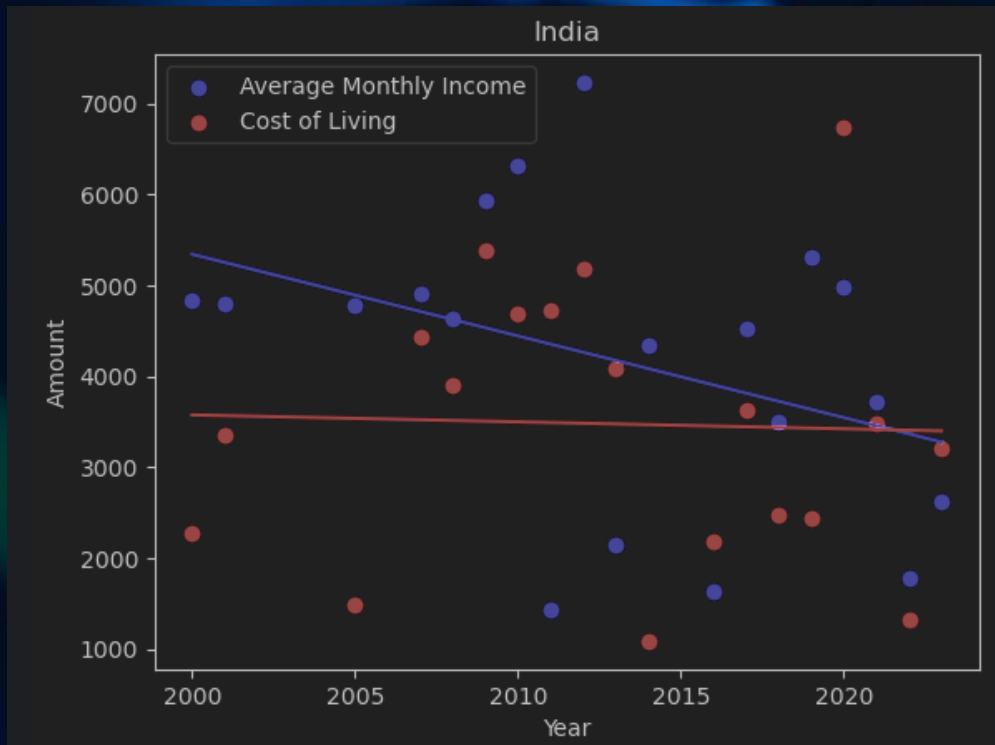
ERGEBNIS

- Canada besitzt besten Verhältnis für die Bevölkerung
- Gefolgt von Australien



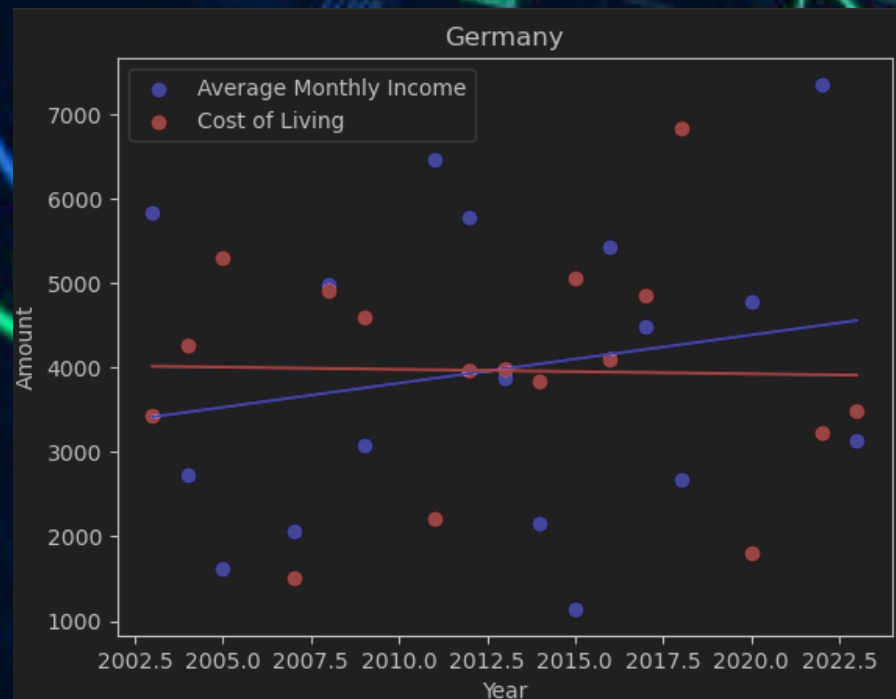
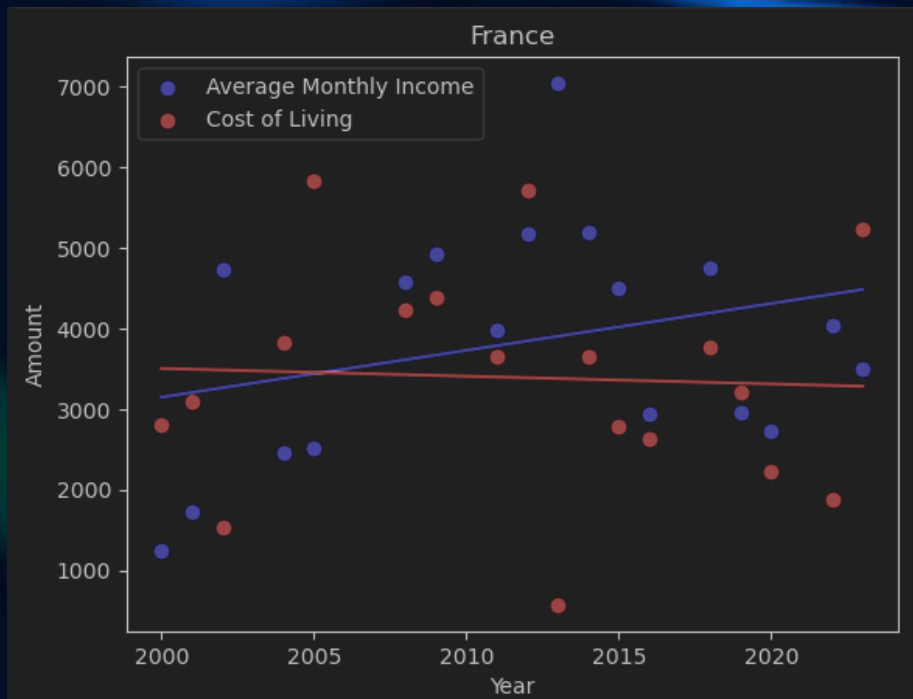
ERGEBNIS

- Schlechtestes Verhältnis: United States und India



ERGEBNIS

- Europa im Vergleich



VIELEN DANK

