Assignment05

Jasmine Liu

2024-10-07

Research Question: Does a higher adult literacy rate or primary school enrolment rate correlate with better GDP per capita?

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv("/Users/jasmineliu/Downloads/QTM350/Assignment05/wdi.csv")

data.head(10)
columns = data.columns.to_list()
print(columns)
print(data["country"].nunique())
##we have 217 countries in our dataset
```

```
['country', 'inflation_rate', 'exports_gdp_share', 'gdp_growth_rate', 'gdp_per_capita', 'adul
217
```

Since we are interested in adult literacy rate, primary school enrolment rate, GDP growth rates, and GDP per capita, let's subset the dataset into only these columns for our analysis.

```
df = data[["country","adult_literacy_rate", "primary_school_enrolment_rate", "gdp_growth_rate

df.head()
```

	country	adult_literacy_rate	primary_school_enrolment_rate	gdp_growth_rate	gdp_per_
0	Afghanistan	NaN	NaN	-6.240172	352.6037
1	Albania	98.5	95.606712	4.856402	6810.114
2	Algeria	NaN	108.343933	3.600000	5023.252
3	American Samoa	NaN	NaN	1.735016	19673.39
4	Andorra	NaN	90.147346	9.563798	42350.69

We see that there's some missing values. Before we proceed to our analysis, let's first preprocess our dataset.

```
df.isna().sum()
cleaned_data = df.dropna(subset=['adult_literacy_rate', 'primary_school_enrolment_rate', 'gd
print(cleaned_data["country"].nunique())
```

```
32
```

It turns out that after dropping all the missing values, we only have 32 countries left to analyze. Considering the small scale of this assignment, let's proceed with these countries.

```
print(cleaned_data["country"].unique())
display(cleaned_data)
```

```
['Albania' 'Bahrain' 'Benin' 'Bhutan' 'Bosnia and Herzegovina'
 'Burkina Faso' 'Burundi' 'Cambodia' 'Chad' 'Chile' 'Dominican Republic'
 'Ecuador' 'Egypt, Arab Rep.' 'Gambia, The' 'Georgia' 'Guatemala' 'India'
 'Kenya' 'Lao PDR' 'Lesotho' 'Madagascar' 'Malawi' 'Morocco' 'Niger'
 'Oman' 'Rwanda' 'Senegal' 'Tanzania' 'Uzbekistan' 'Viet Nam'
 'West Bank and Gaza' 'Zimbabwe']
```

	country	adult_literacy_rate	primary_school_enrolment_rate	gdp_growth_rate
1	Albania	98.500000	95.606712	4.856402
14	Bahrain	97.872482	92.344193	4.891500
20	Benin	47.099998	113.048912	6.253245
22	Bhutan	72.099998	103.800003	5.213868
24	Bosnia and Herzegovina	98.300003	87.822220	4.226811
30	Burkina Faso	34.490002	82.356796	1.777915
31	Burundi	75.540001	103.901001	1.848999
33	Cambodia	83.779999	109.959000	5.239810

	country	adult_literacy_rate	primary_school_enrolment_rate	gdp_growth_rate
38	Chad	27.280001	90.364120	2.804341
40	Chile	97.160004	100.192268	2.058740
56	Dominican Republic	95.500000	100.492287	4.858324
57	Ecuador	93.948120	97.534103	6.186139
58	Egypt, Arab Rep.	74.500000	91.590179	6.587846
71	Gambia, The	58.669998	92.320000	4.917821
72	Georgia	99.574989	104.548119	10.958532
80	Guatemala	84.269997	103.906227	4.118004
89	India	76.322777	111.084000	6.987039
101	Kenya	82.879997	97.185997	4.846635
108	Lao PDR	87.519997	97.223000	2.707427
111	Lesotho	82.010002	88.611778	1.285504
118	Madagascar	77.480003	138.192001	3.796199
119	Malawi	68.080002	126.438004	0.923450
133	Morocco	77.349998	114.169991	1.258544
143	Niger	38.099998	68.331413	11.900000
148	Oman	97.339058	90.096809	4.311254
162	Rwanda	78.763184	134.930344	8.157734
167	Senegal	57.669998	83.278702	3.821856
191	Tanzania	82.019997	95.504997	4.566006
208	Uzbekistan	99.999977	94.194000	5.668817
211	Viet Nam	96.133263	123.134003	8.123514
213	West Bank and Gaza	97.843842	91.764587	4.082760
216	Zimbabwe	89.849998	95.790001	6.522375

The countries we have: 'Albania', 'Bahrain', 'Benin', 'Bhutan', 'Bosnia and Herzegovina', 'Burkina Faso', 'Burundi', 'Cambodia', 'Chad', 'Chile', 'Dominican Republic', 'Ecuador', 'Egypt, Arab Rep.', 'Gambia, The', 'Georgia', 'Guatemala', 'India', 'Kenya', 'Lao PDR', 'Lesotho', 'Madagascar', 'Malawi', 'Morocco', 'Niger', 'Oman', 'Rwanda', 'Senegal', 'Tanzania', 'Uzbekistan', 'Viet Nam', 'West Bank and Gaza', 'Zimbabwe'

This means that the analysis we conduct will only be applicable to these countries.

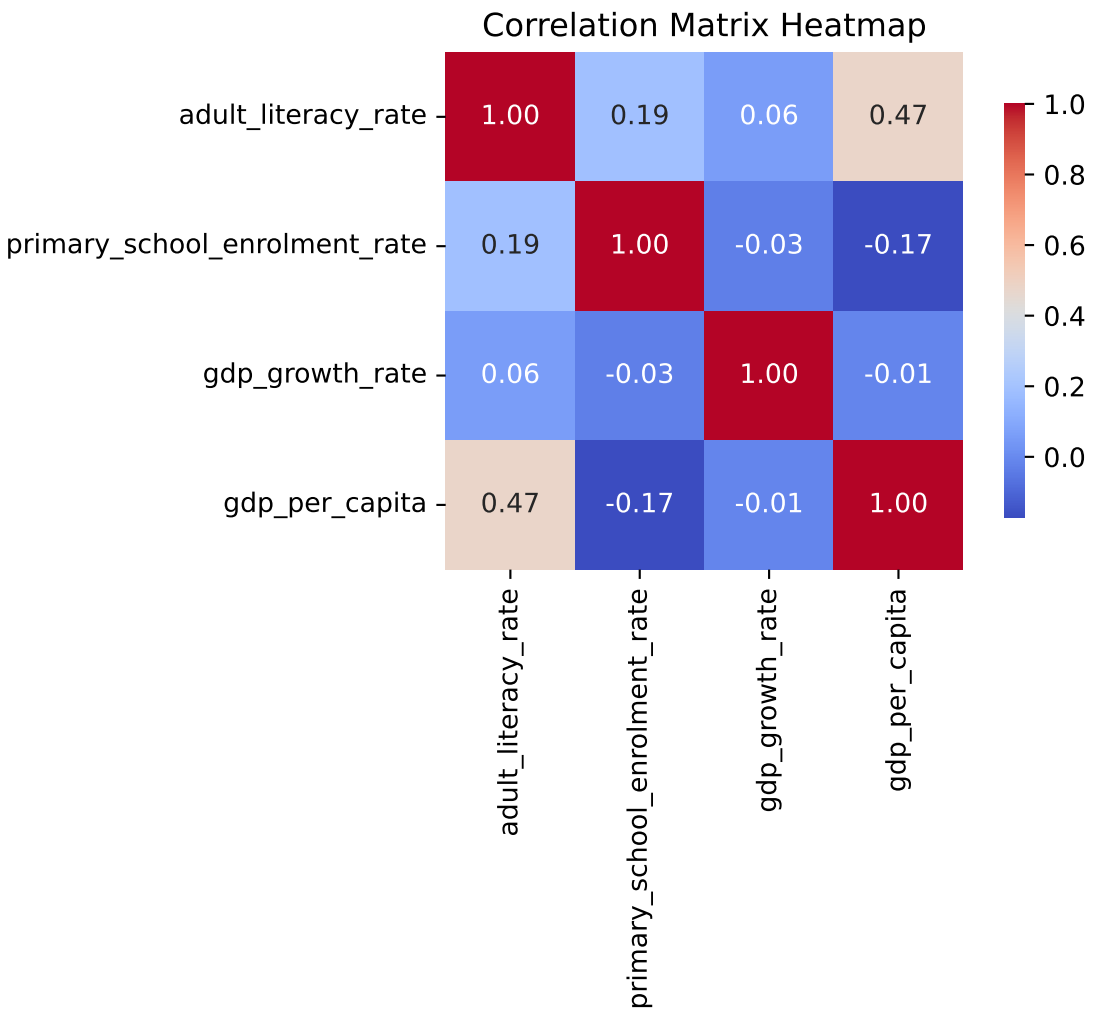
Here's descriptive analysis that provides us an idea of the features' mean values, IQR values, and min, max values!

```
display(cleaned_data['adult_literacy_rate'].describe())
display(cleaned_data['primary_school_enrolment_rate'].describe())
display(cleaned_data['gdp_growth_rate'].describe())
display(cleaned_data['gdp_per_capita'].describe())
```

```
count    32.000000
mean     78.998365
std      20.067613
min      27.280001
25%      73.900000
50%      82.449997
75%      96.389948
max      99.999977
Name: adult_literacy_rate, dtype: float64

count     32.000000
mean     100.616118
std       15.017442
min       68.331413
25%       91.720985
50%       97.204498
75%      105.900839
max      138.192001
Name: primary_school_enrolment_rate, dtype: float64

count    32.000000
mean      4.867419
std       2.564314
min       0.923450
25%       3.548235
50%       4.851518
75%       6.202916
max      11.900000
Name: gdp_growth_rate, dtype: float64

count       32.000000
mean      4870.336536
std       6839.795005
min        259.025031
25%        986.285820
50%       2187.729901
75%       5724.067567
max      30146.925026
Name: gdp_per_capita, dtype: float64
```

Now, let's do some analysis!

```
correlation_matrix = cleaned_data[['adult_literacy_rate', 'primary_school_enrolment_rate', 'g

sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap='coolwarm', square=True, cbar_kws
plt.title('Correlation Matrix Heatmap')
plt.show()
```



We see that the correlation between GDP per capita and adult literacy rate is 0.47, let's see what they look like in a scatterplot!

We can see a pretty clear exponential curve that as adult literacy rate increases, the GDP per capita tends to increase as the adult literacy rate reaches above 70%!

Unlike the previous figure, we can't really see a general trend between primary school enrolment rate vs gdp per capita.

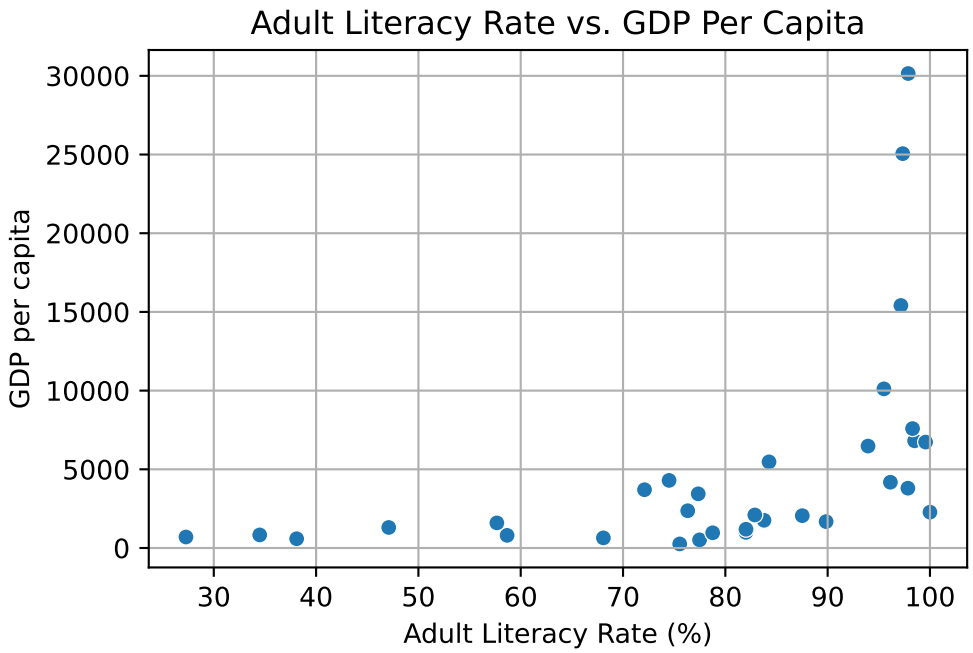


Figure 1

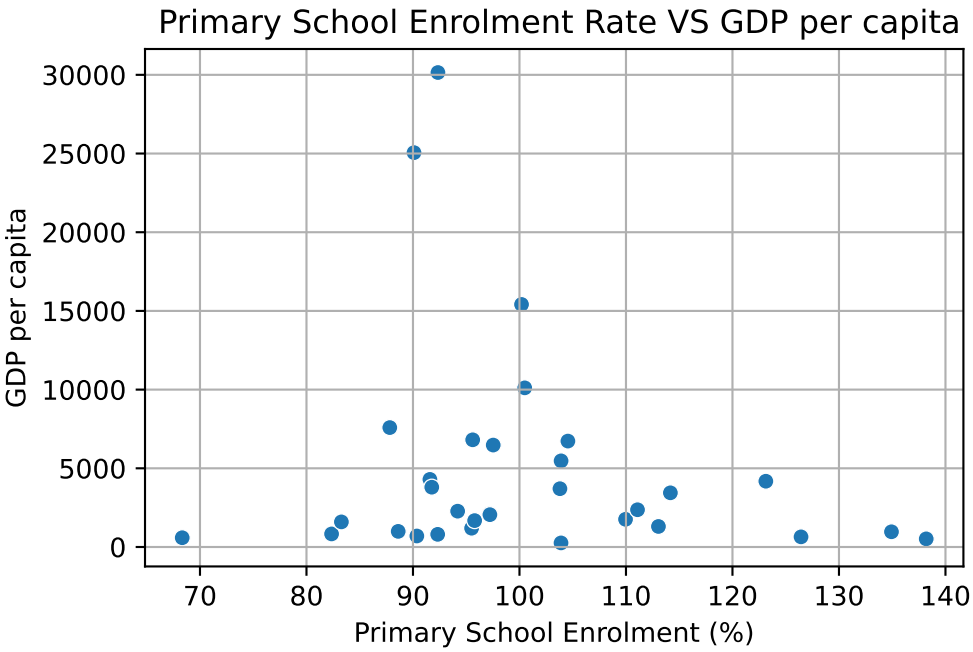


Figure 2

Let's construct a bar graph that classify the adult literacy rate into low, medium, high, and very high VS GDP per capita!

```
/var/folders/sn/6ghc2n6d7qd3q_hb0b_9_cnr0000gn/T/ipykernel_5628/2009248937.py:1: SettingWithC


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide

/var/folders/sn/6ghc2n6d7qd3q_hb0b_9_cnr0000gn/T/ipykernel_5628/2009248937.py:6: FutureWarnin


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assig
```

Based on Figure 1 and Figure 3, there is a noticeable correlation between adult literacy rates and GDP per capita, specifically telling us that higher literacy rate is associated with higher GDP per capita. This finding aligns with other research. According to Dr. Rahman from the University of Chittagong, "taking proper initiative to increase literacy rate of a country will reduce its unemployment rate and increase PGDP resulting development of the country" [**?**]. Additionally, the World Literacy Foundation highlights the importance of education in economic development as "the cost of illiteracy to the global economy is estimated at USD $1.19 trillion" in 2022 [**?**].

	country	adult_literacy_rate	primary_school_enrolment_rate	gdp_gr
1	Albania	98.500000	95.606712	4.856
2	Bahrain	97.872482	92.344193	4.891
3	Benin	47.099998	113.048912	6.253
4	Bhutan	72.099998	103.800003	5.213
5	Bosnia and Herzegovina	98.300003	87.822220	4.226
6	Burkina Faso	34.490002	82.356796	1.777
7	Burundi	75.540001	103.901001	1.848
8	Cambodia	83.779999	109.959000	5.239
9	Chad	27.280001	90.364120	2.804
10	Chile	97.160004	100.192268	2.058
11	Dominican Republic	95.500000	100.492287	4.858
12	Ecuador	93.948120	97.534103	6.186
13	Egypt, Arab Rep.	74.500000	91.590179	6.587

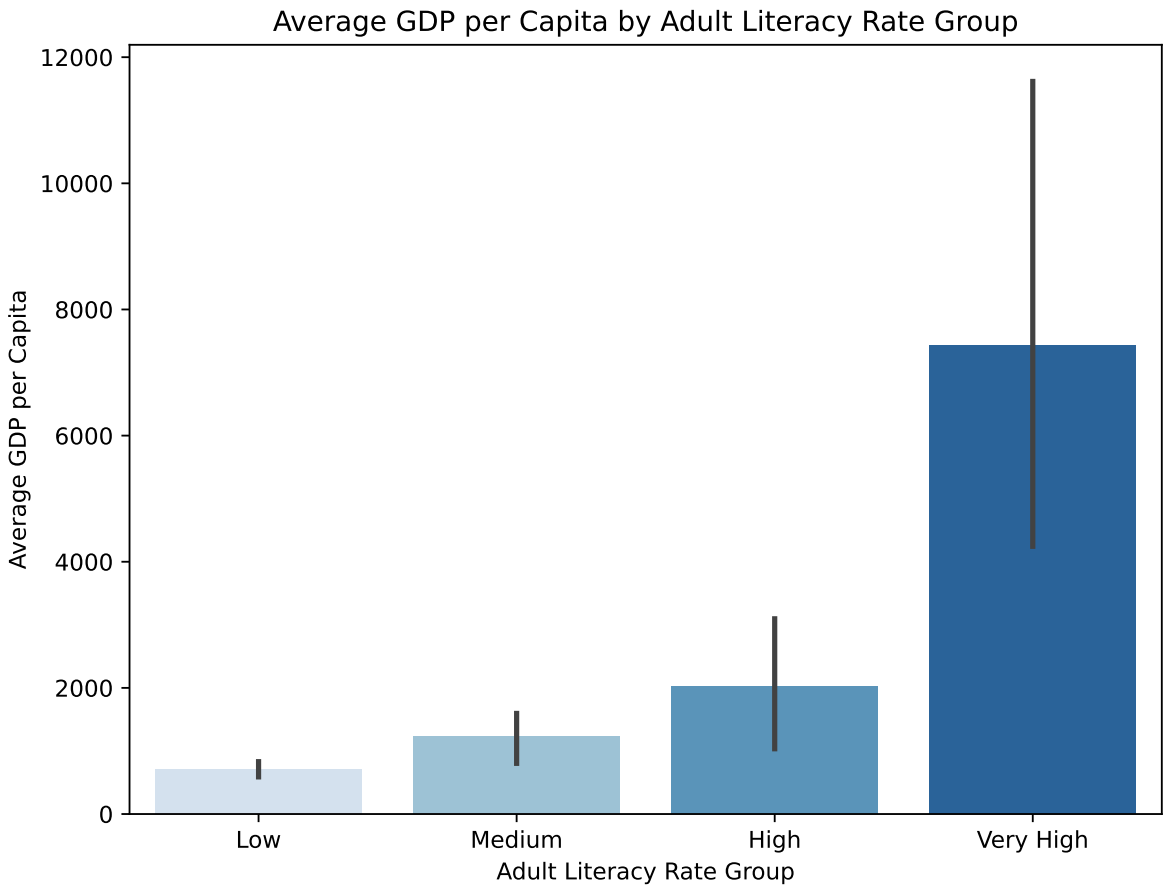

Average GDP per Capita by Adult Literacy Rate Group
Average GDP per Capita
12000
10000
8000
6000
4000
2000
0
Low
Medium
High
Very High
Adult Literacy Rate Group

Figure 3

14	Gambia, The	58.669998	92.320000	4.9178
15	Georgia	99.574989	104.548119	10.958
16	Guatemala	84.269997	103.906227	4.1180
17	India	76.322777	111.084000	6.9870
18	Kenya	82.879997	97.185997	4.8466
19	Lao PDR	87.519997	97.223000	2.7074
20	Lesotho	82.010002	88.611778	1.2855
21	Madagascar	77.480003	138.192001	3.7961
22	Malawi	68.080002	126.438004	0.9234
23	Morocco	77.349998	114.169991	1.2585
24	Niger	38.099998	68.331413	11.900
25	Oman	97.339058	90.096809	4.3111
26	Rwanda	78.763184	134.930344	8.1577
27	Senegal	57.669998	83.278702	3.8218
28	Tanzania	82.019997	95.504997	4.5660
29	Uzbekistan	99.999977	94.194000	5.6688
30	Viet Nam	96.133263	123.134003	8.1235
31	West Bank and Gaza	97.843842	91.764587	4.0827
32	Zimbabwe	89.849998	95.790001	6.5223