Leitfaden für nachvollziehbare Schritte

1. Kurze Darstellung des Problembereichs / Aufriss des Themas

1.1 Inhaltlich

Kern der Untersuchung : Data analysis

Grobziele der Arbeit : Impact of COVID-19 on education globally.

In this project we have the data analysis on the socio economic status of arround 200 countries on education during the COVID-19 pandemic.

1.2 Begründung desThemas

Darstellung der Relevanz des Themas?

Warum ist das Thema wichtig und interessant und daher bearbeitungs- und förderungswürdig?

School closures due to pandamic have brought significant disruption to education.

Which is indirectly effecting the socio economic status of the country and the world alltogether.

The pandemic has created learning losess in children and an undefined gap in education system . which has to be filled with differerent stratergic plans and analysis .

Darstellung eines persönlichen Erkenntnisinteresses.

Dieser Abschnitt soll ein prägnanter Einstieg in die Projektarbeit / Seminararbeit sein.

Er soll beim Leser Interesse für das Thema und die Bereitschaft wecken oder verstärken, die Arbeit zu betreuen bzw. zu fördern und dient der Eigenmotivation.

The pandemic brought the educatioin system of whole world to halt. For some was just a school closure for few months or an year but for the some children was never return to the school in their life time.The loss of learning is unaccetable morally.The impact is high interms of productivty ,earnings, well being for this generation of children and the world's economy. To recover from the loss countries should build strong education system which would be unaffected by future closures .Increasing the resources interms of technology to reach the learners in an effective way.

2. Nachvollziehbare Schritte

 Detailed analysis exits on many internet sites including wikipedia

https://en.unesco.org/covid19/educationresponse

https://en.wikipedia.org/wiki/Impact_of_the_COVID-19_pandemic_on_education

https://www.worldbank.org/en/news/press-release/2021/12/06/learning-losses-from-covid-19-could-cost-this-generation-of-students-close-to-17-trillion-in-lifetime-earnings

2.1 Der Stand der Forschung / Auswertung der vorhandenen Literatur / Tutorials ...

Wurde das Problem früher bereits untersucht?

Welche Aspekte wurden untersucht und welche nicht?

Welche Kontroversen gab es und welche Methoden standen bis jetzt im Vordergrund?

Yes, there are analysis done by well known oroganizations such as UNESCO, World Bank

Lösungswege strukturieren!

Importing important libraries

Load the dataset into a data frame using Pandas

Explore the number of rows & columns, ranges of values etc.

Handle missing, incorrect and invalid data

Performing any additional steps

Compute the median for numeric columns

Explore distributions of numeric columns using bar graphs

Explore relationship between columns using scatter plots,stacked plots .

Wichtigste (verwendete) wissenschaftliche Positionen zum ausgewählten Thema?

(Z.B. **Tutorials …**)

2.2 Fragestellung

Is pandemic impacts is same in developed and developing world?
Has the pandemic created any learning losses throughout the World?

2.3 Stand der Forschung

Due to high impact of COVID-19 pandemic on education there is need for deatailed

Analysis and measure to over come the learning loss.

2.4 Wissenslücke

Some important ascpects which could afftect education were not considered during data collection

2.5 Methode

Detaillierte nachvollziehbare Beschreibung der Vorgehensweise !!

1) Importing the mail libraries Numpy,pandas,matplotlib,seaborn and datetime

2) Import the main data record as a data frame (education_COVID-19__)

3) Delete unnecessary columns

4) Replace nan values with meadian in numerical columns

```python
1   # -*- coding: utf-8 -*-
2   """
3   Created on Thu Jan 27 11:59:00 2022
4
5   @author: Vaishu
6   """
7   import numpy as np
8   import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
9   import matplotlib.pyplot as plt
10  import seaborn as sns
11  import datetime
12  #Reading the csv file in pandas
13
14  df=pd.read_csv(r"C:\Users\Vaishu\Desktop\Work\project_01\education_COVID-19__.csv")
15  print(df.head(200))
16  # Deatils of the data frame
17  print(df.shape)
18  print(df.describe())
19  print(df.columns)
20  #Dealing with the date format
21  df['If Closed Due To COVID19 When'] = pd.to_datetime(df['If Closed Due To COVID19 When'], errors
22  df['If Closed Due To COVID19 When'] = df['If Closed Due To COVID19 When'].replace(np.nan, 0)
23  print(df)
24
25  # Replacing commas with ''
26
27  df1=df.replace(",","",regex=True) #converting comma(,) to ()
28
29
30
```

Checking on the Data Types and Deleting the columns

```python
30
31  #%%
32  # checking on data type
33  for column in df1.columns:
34      print(df1[column].dtype)
35
36  #Deleting some columns
37  df2 = df1.drop(labels=["Latitude", "Longitude","Year Pre", "Year Prm", "Year Sec", "Year Ter"], axis=1)
38  print(df2)
39
40  print(df2.columns)
41
```

Converting Non numerical Value to numerica values

```
41
42  #%%
43  #converting non numeric values to numeric values
44
45  df2[["Enrollment", "Se Pre Enrl","Se Sec Enrl","Se Prm Enrl","Se Ter Enrl"]] = df2[["Enrollment", "Se Pre Enrl","
46
47  #replacing NaN values to median
48
49  df3=df2.fillna({'Enrollment': df2['Enrollment'].median(),
50              'Se Pre Enrl': df2['Se Pre Enrl'].median(),
51              'Se Sec Enrl': df2['Se Sec Enrl'].median(),
52              'Se Prm Enrl': df2['Se Prm Enrl'].median(),
53              'Se Ter Enrl': df2['Se Ter Enrl'].median()})
54
55
56
```

Printing Unique values in all the columns

```
56
57  #%%
58  #printing unique values in all the column
59
60  for values in df3:
61      print("{} has {} values.They are:".format(values,len(df3[values].unique())))
62      print(df3[values].unique())
63      print('\n')
64
```

Handling the Missing Data

```
64
65  #%% getting number of missing data in each column
66  print(df3.isna().sum())
67  # Handling the missing data
68  index_with_nan = df3.index[df3.isnull().any(axis=1)]
69  print( index_with_nan)
70  print(df3)
```

Filling the Missing data

```
70  print(df3)
71  #%%
72  #Filling the missing numerical data with median
73
74  df4=df3.replace(",","",regex=True)
75  df4[['Enrollment', 'Se Pre Enrl', 'Se Prm Enrl','Se Sec Enrl', 'Se Ter Enrl']]=df4[['Enrollment', 'Se Pre Enrl',
76
77  df4=df3.fillna(df3.median())
78  print(df4)
79  #Renaming the columns in df4
80  df4.rename(columns={'Country Name':'country_name','Se Pre Enrl': 'Pre_enrl', 'Se Prm Enrl': 'Prm_enrl','Se Sec En
81
82  print(df4)
83  print(df4.columns)
```

Ploting the graph of Preprimary Enrollment in 50 countries

```
87
88  #%%
89  #ploting the graph country vs  enrollments in pri primary uning pandas
90
91  df4.iloc[:50].plot.bar(x='country_name', y='Pre_enrl')
92  plt.rcParams["figure.dpi"]=120
93  plot=df4.plot.bar(x='country_name', y='Pre_enrl',figsize=(40,10))
94  fig=plot.get_figure()
95  fig.savefig("output1.png")
96
```
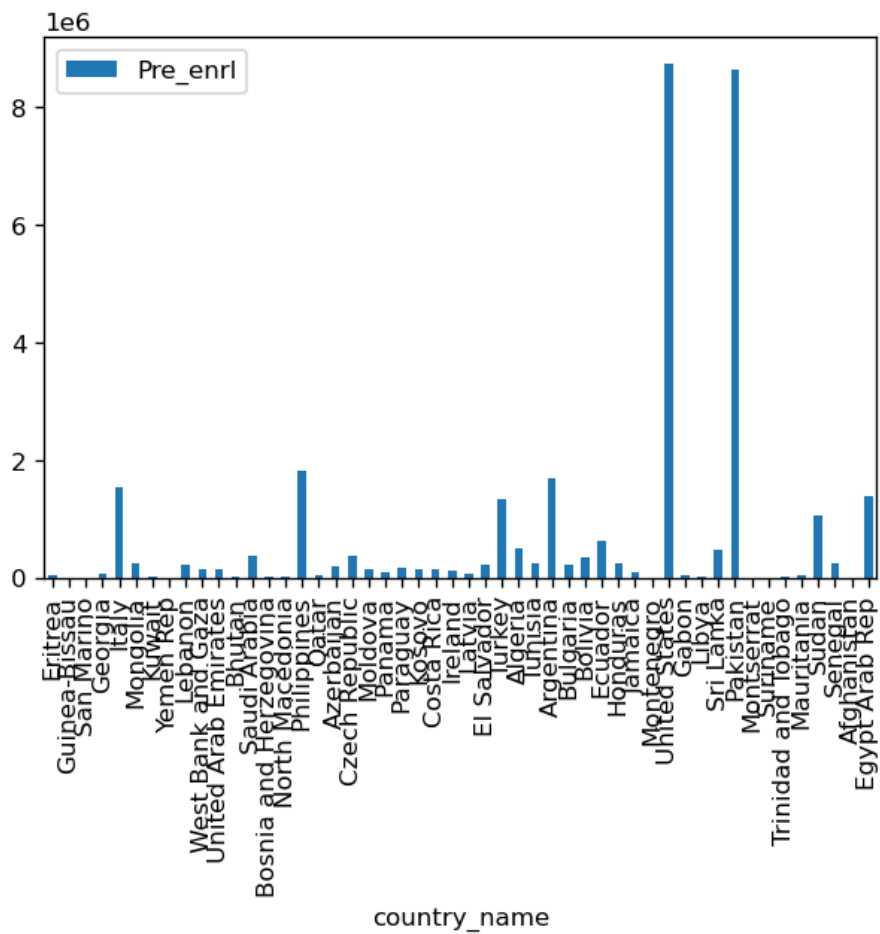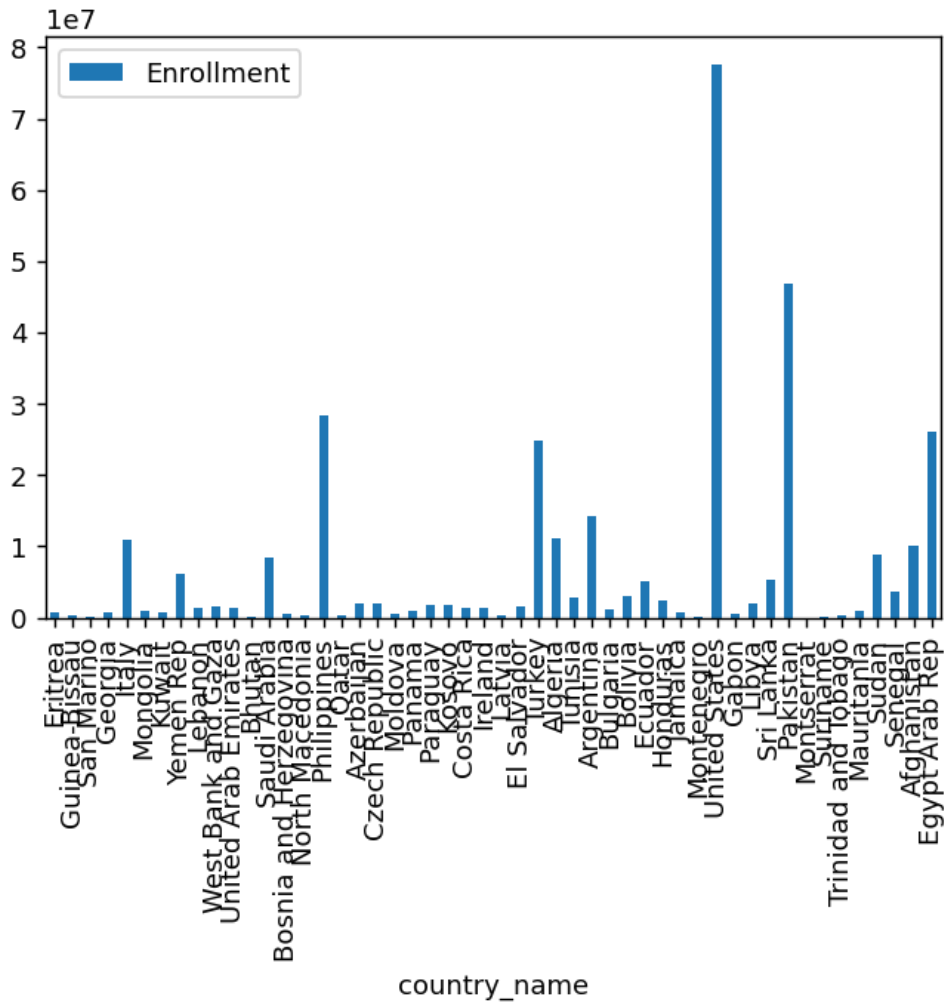
Ploting the graph of total number of Enrollment in 50 countries

```
101  #%%
102  #ploting the graph country vs total number of enrollments  uning pandas
103
104  df4.iloc[:50].plot.bar(x='country_name', y='Enrollment')
105  plt.rcParams["figure.dpi"]=120
106  plot=df4.plot.bar(x='country_name', y='Enrollment',figsize=(50,30))
107  fig=plot.get_figure()
108  fig.savefig("output2.png")
100  #%%
```
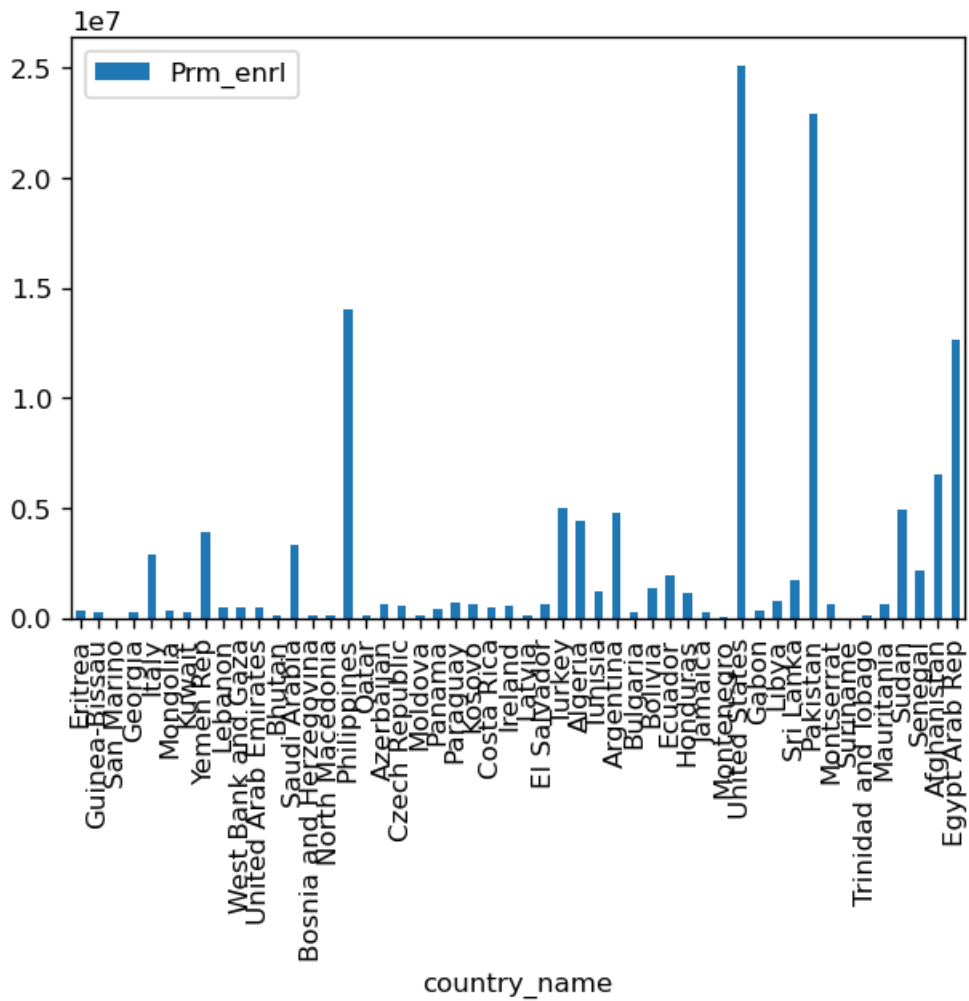
Ploting the graph of Primary Enrollment in 50 countries

```
109  #%%
110
111  #ploting the graph country vs  enrollments in primary uning pandas
112
113  df4.iloc[:50].plot.bar(x='country_name', y='Prm_enrl')
114  plt.rcParams["figure.dpi"]=120
115  plot=df4.plot.bar(x='country_name', y='Prm_enrl',figsize=(40,10))
116  fig=plot.get_figure()
117  fig.savefig("output3.png")
```
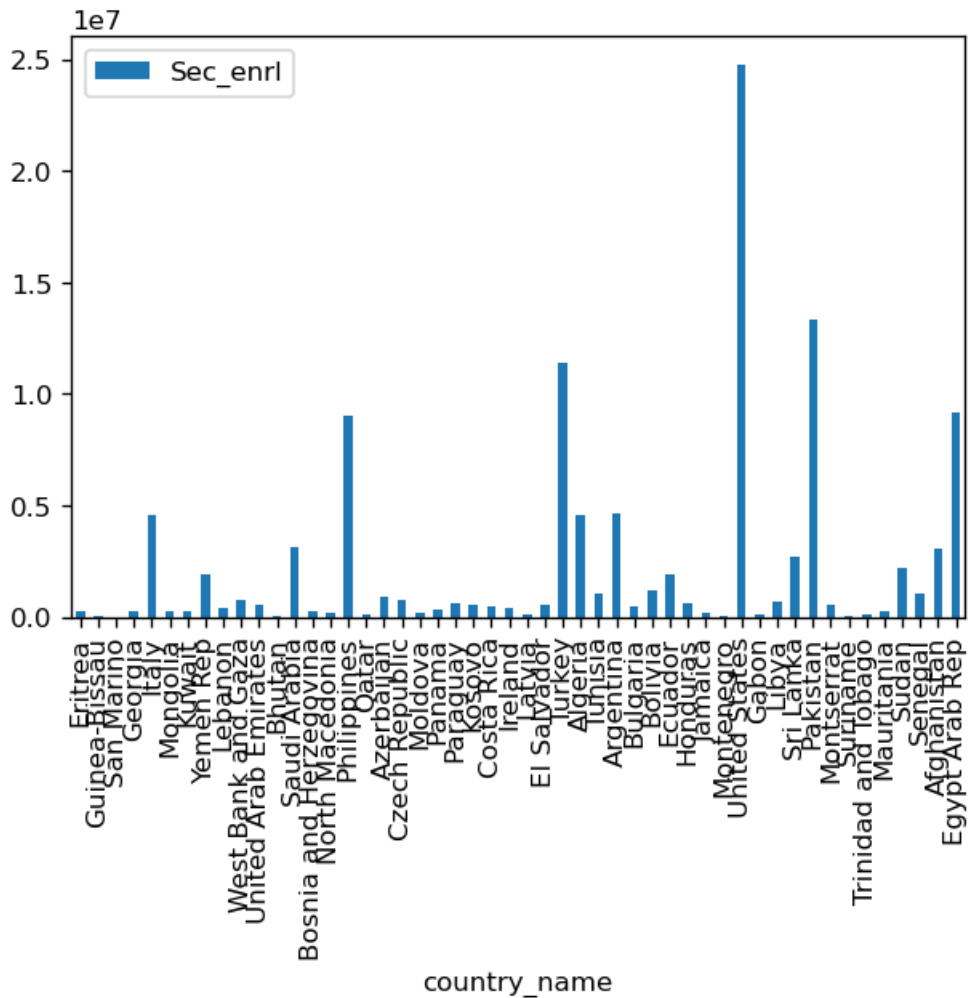
**Ploting the graph of secondary Enrollment in 50 countries**

```
118  #%%
119  #ploting the graph country vs  enrollments in secondary uning pandas
120
121  df4.iloc[:50].plot.bar(x='country_name', y='Sec_enrl')
122  plt.rcParams["figure.dpi"]=120
123  plot=df4.plot.bar(x='country_name', y='Sec_enrl',figsize=(40,10))
124  fig=plot.get_figure()
125  fig.savefig("output4.png")
```

Ploting the graph of tertiary Enrollment in 50 countries

```
126  #%%
127  #ploting the graph country vs  enrollments in tertiary uning pandas
128
129  df4.iloc[:50].plot.bar(x='country_name', y='Ter_enrl')
130  plt.rcParams["figure.dpi"]=120
131  plot=df4.plot.bar(x='country_name', y='Ter_enrl',figsize=(40,10))
132  fig=plot.get_figure()
133  fig.savefig("output5.png")
134
135
```
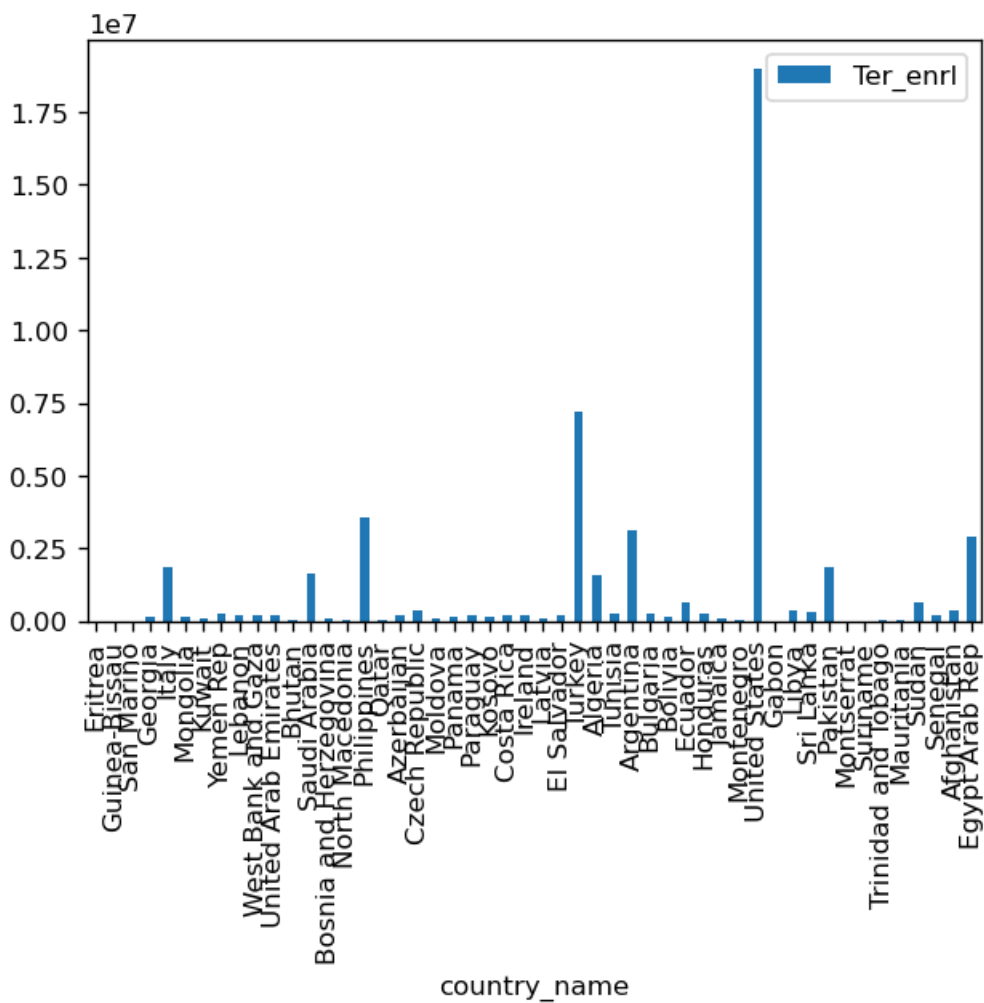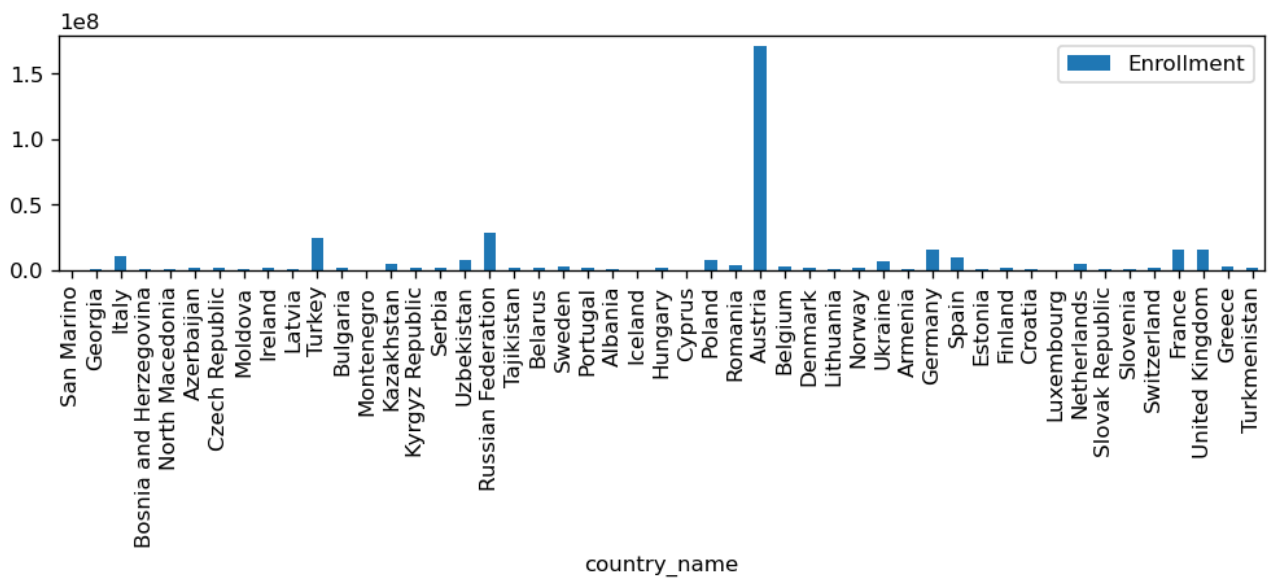

Ergebnis

The graphs above clearly shows there were good number of enrollments in United States and Pakistan in Preprimary ,Primary and Secondary levels.While countries like Philippines and Egypt Arab

have quite a good number of enrollments in secondary level. Comparing in Tertiary level only United States have high number of enrollments.

Comparision of Enrollment only in the Region Europe and Central Asia

```
136  #%% study on enrollements in Europe and Central Asia
137
138  df_europe = df4[df4["Region Name"]=="Europe and Central Asia"][["Enrollment","country_name"]]
139  print(df_europe)
140  print(df_europe.describe())
141  print(df_europe.shape)
142
143  plot=df_europe.plot.bar(x='country_name', y='Enrollment',figsize=(10,2))
144
145
```
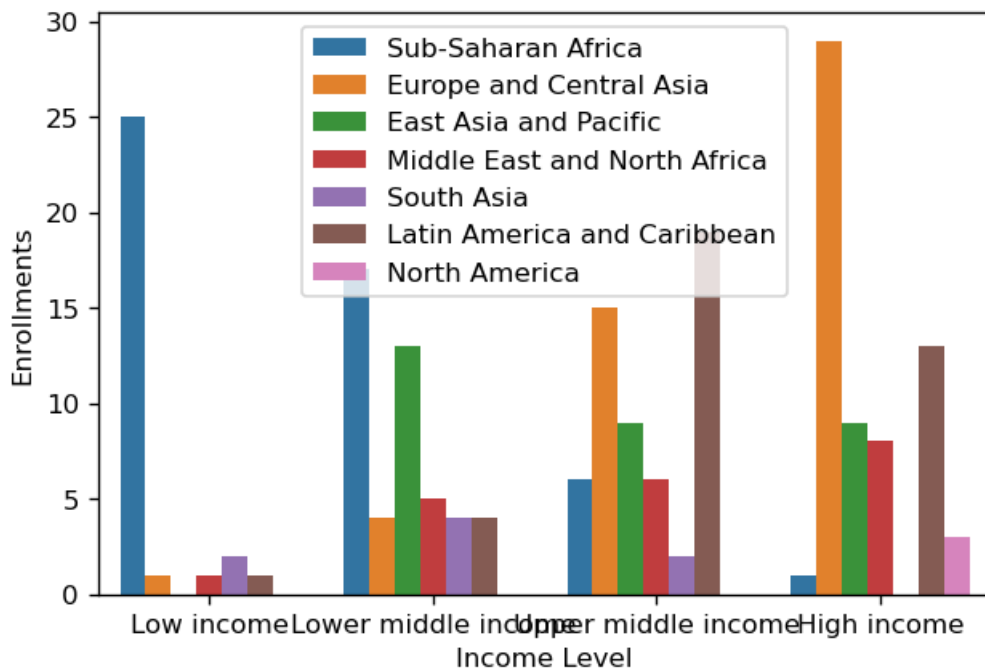

Ergebnis

The graph above clearly shows only Austria had high number of enrolmments compared to other in the region Europe and Central Asia.

Comparing Enrollments grouping them in income level

```
146  #%%
147  #comparing Income level with Enrollment
148
149  ax=sns.countplot(x="Income Level",hue="Region Name",data=df4,order=["Low income","Lower middle income","Upper mid
150  plt.legend(loc="best", frameon=True)
151  plt.show()
152
```
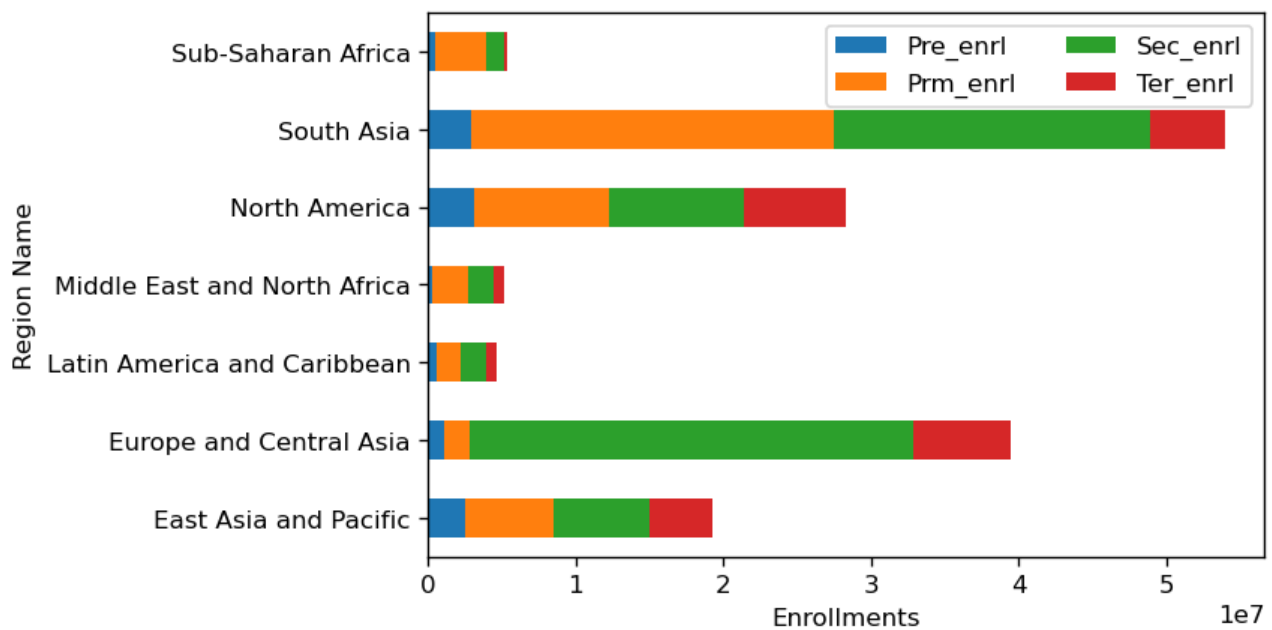

Ergebnis

The graph above gives an over view of enrollments in different region when catagerized on the basis of income level. Sub-Saharan countries with low income and lower middle income had good number of enrollments .

Europe and central Asian couries with high Income also had high number of enrollments.

East Asia and Pacific countries with low income had no enrollments compared to the countries with other income level in that region.The countries in the region Latin America and Caribbean had good number of enrollments in uppermiddle incom and high income compared with other income catagory.The countries in the region North America had very lowest enrollments.

Stacked plot showing Enrollments in accordance with region

```
175  sns.relplot(x= country_name , y= Enrollment , kind= scatter , data=df4);
176  #%%
177  #split Apply Combine
178  #Grouped by Region name
179  g = df4.groupby("Region Name")
180  print(g)
181
182
183  for region, y in g:
184      print("Region Name:",region)
185      print("\n")
186      print("data:",y)
187  df_region=g.mean()
188  print(df_region)
```

```
192  #%%
193  df_region[['Pre_enrl','Prm_enrl','Sec_enrl','Ter_enrl']].plot.barh(stacked =True)
194  plt.legend(ncol=2, loc="best", frameon=True)
195  plt.xlabel("Enrollments")
196  plt.show()
197
```


Ergebnis

The stacked graph above shows Enrollments in different levels of education sector in different regions

It clearly shows South Asia had good number of enrollments in Primary and Secondary education.

Europe and Central Asia had high number of enrollment in Secondary education compare other.

Sub-saharan Africa, Middle East and North Africa and Latin Anrica and caribbean had all together very less number of enrollment in sections of education

Stacked plot showing Toltal no. Of Enrollments in accordance with level of income

```python
#%%
i=df4.groupby("Income Level")
print(i)


for income, x in i:
    print("Income Level:",income)
    print("\n")
    print("data:",x)
df_income=i.mean()
print(df_income)


#%% comparision of Enrollments ion the basis of Income

df_income[['Pre_enrl','Prm_enrl','Sec_enrl','Ter_enrl']].plot.barh(stacked =True)
plt.legend(ncol=2, loc="best", frameon=True)
plt.xlabel("Enrollments")
plt.show()
```
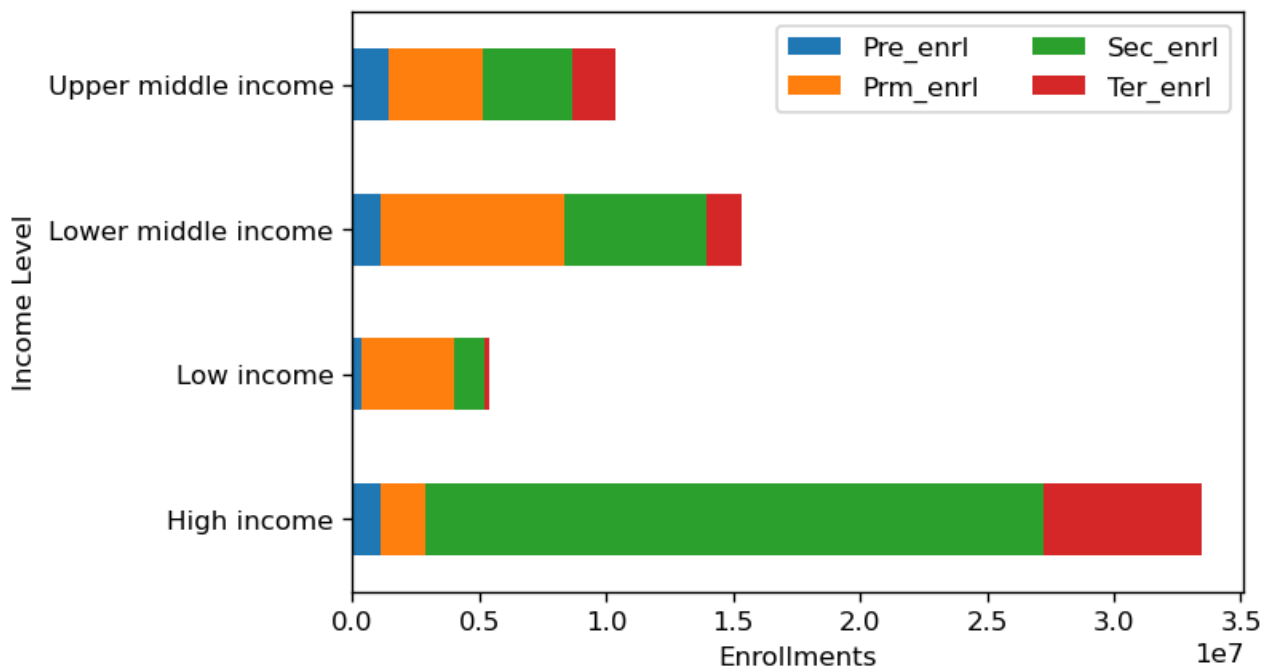

Ergebnis

The above graph shows comparison between number of enrollments with Income level

The countries with high income level have high enrollments in secondary education compare to other sections.The lower middle income couries ha almost equal number of enrollments in Primary and Secondary education .But in all the levels the number of enrollment are very less in Tertiary and preprimary sections.

scatter plot showing closure of schools thoughout the world

```
218
219  #%%
220  #comparision of status of the schools in different region during the pandemic
221
222  df4['If Closed Due To COVID19 When']=df4['If Closed Due To COVID19 When'].astype('string')
223
224  df4['COUNTER'] =1
225  df_status = df4.groupby(["If Closed Due To COVID19 When","School Status","Region Name"])['COUNTER'].sum() #sum fu
226  print(df_status)
227
228
229
230  fig = plt.gcf()
231
232  # Change seaborn plot size
233  fig.set_size_inches(15, 4)
234  sns.scatterplot(data=df_status, x="If Closed Due To COVID19 When", y="Region Name",hue="School Status",s=80)
235  plt.legend(loc="lower right", frameon=True)
236  plt.rcParams["figure.dpi"]=120
237  plt.xticks(rotation =90)
238  plt.show()
239
240
```
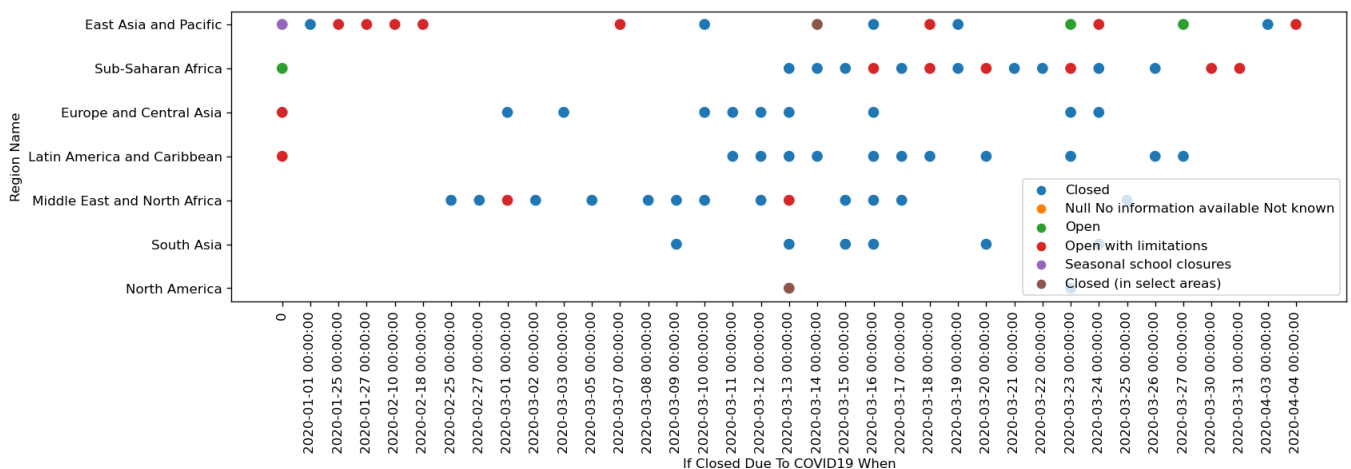

Ergebnis

This scatter plot clearly indiactes that during the Pandemic most of the schools were closed in all regions/countries

2.7   Ausblick

With above data set the impact of COVID-19 can be analysed in many othe directions .

A systemic plan and action could taken to over come the loss we had in our Education system.