

# Merging notes

## 1. GIS data

### first principle is country in pathway data.

1. pathway data with population data:

Changed the country name

```
'United Republic of Tanzania':'Tanzania',  
'Republic of Moldova': 'Moldova',  
'Bolivia (Plurinational State of)':'Bolivia'
```

2. pathway data with NEET data

All of 97 countries in datapath data are in NEET data.

Using NEET data country as first principle and set missed GIS data as nan.

## 2. Return model

### first principle is country in pathway data.

1. GIS data with year of duration for each school level

Taken the latest valid number for each country (224 areas in total)

Left merge with GIS data

Fill nan by – primary : 6 years, LS : 3 years, US : 3 years

2. GIS data with GDP values (Dollar per Captial)

Taken the latest valid number for each country (265 areas in total)

change the country name

```
'Egypt, Arab Rep.': 'Egypt', 'Congo, Dem. Rep.': 'Democratic Republic of the C  
'Congo, Rep.': 'Congo',  
'Lao PDR': 'Lao People\'s Democratic Republic',  
'St. Lucia': 'Saint Lucia',  
'Yemen, Rep.' : 'Yemen'
```

Missing data as nan/zeros

3. GIS data with return rates from the paper

Drop empty area. 93 area remains.

Filling the missed data as 10% increase rate.

## 3. Benefit model

1. Return model with incoming categories data.

change the country name in income categories data,

```
'Congo, Rep.': 'Congo',  
'Congo, Dem. Rep.': 'Democratic Republic of the Congo',  
'Egypt, Arab Rep.': 'Egypt',  
'Lao PDR': 'Lao People\'s Democratic Republic',  
'St. Lucia': 'Saint Lucia',  
'São Tomé and Príncipe': 'Sao Tome and Principe',  
'Yemen, Rep.': 'Yemen'
```

Missing data of Kyrgyzstan and State of Palestine are filled – lower middle income, sou

2. Return model with HLO data

Using the 2020 data(latest)

Left merge and fill the missing HLO as the mean of each income group.