education_information

May 1, 2022

Education Data

[1]:
```
import pandas as pd
import numpy as np

df_education = pd.read_csv ('../Data/Education/Education_Data_1970.csv')

df_education.head()
```

[1]:
```
  Country Name Country Code  \
0  Afghanistan          AFG
1  Afghanistan          AFG
2  Afghanistan          AFG
3  Afghanistan          AFG
4      Albania          ALB

                                              Series          Series Code  \
0  Government expenditure on education as % of GD...    SE.XPD.TOTL.GD.ZS
1  Expenditure on education as % of total governm...    SE.XPD.TOTL.GB.ZS
2  Government expenditure on primary education, P...    UIS.X.PPP.1.FSGOV
3  Government expenditure on secondary education,...  UIS.X.PPP.2T3.FSGOV
4  Government expenditure on education as % of GD...    SE.XPD.TOTL.GD.ZS

  1970 [YR1970] 1971 [YR1971] 1972 [YR1972] 1973 [YR1973] 1974 [YR1974]  \
0            ..       1.16036       1.11718       1.42788            ..
1            ..            ..            ..            ..            ..
2            ..            ..            ..            ..            ..
3            ..            ..            ..            ..            ..
4            ..            ..            ..            ..            ..

  1975 [YR1975]  ... 2011 [YR2011] 2012 [YR2012] 2013 [YR2013] 2014 [YR2014]  \
0       1.30332  ...       3.46201        2.6042       3.45446       3.69522
1            ..  ...      16.04843       10.3568       14.1028      14.46593
2            ..  ...    1098.19492     918.52758    1348.24646    1508.31466
3            ..  ...     471.78285     394.59807      579.2046     647.96967
4            ..  ...            ..            ..        3.5393            ..

  2015 [YR2015] 2016 [YR2016] 2017 [YR2017] 2018 [YR2018] 2019 [YR2019]  \
0        3.2558       4.22836       4.05887            ..            ..
```

```
1        12.509       16.2117      15.66138            ..            ..
2     1325.93552    1379.27045     1433.9328            ..            ..
3      569.61989     695.46008      722.1059            ..            ..
4        3.43797       3.96209       3.61172            ..            ..

  2020 [YR2020]
0            ..
1            ..
2            ..
3            ..
4            ..

[5 rows x 55 columns]
```

[2]:
```
df_edu_pri_ppp = df_education[df_education["Series Code"] == 'UIS.X.PPP.1.
 ↪FSGOV']
df_edu_sec_ppp = df_education[df_education["Series Code"] == 'UIS.X.PPP.2T3.
 ↪FSGOV']


df_edu_sec_ppp.head()
```

[2]:
```
      Country Name Country Code  \
3      Afghanistan          AFG
7          Albania          ALB
11         Algeria          DZA
15  American Samoa          ASM
19         Andorra          AND

                                                Series          Series Code  \
3   Government expenditure on secondary education,...  UIS.X.PPP.2T3.FSGOV
7   Government expenditure on secondary education,...  UIS.X.PPP.2T3.FSGOV
11  Government expenditure on secondary education,...  UIS.X.PPP.2T3.FSGOV
15  Government expenditure on secondary education,...  UIS.X.PPP.2T3.FSGOV
19  Government expenditure on secondary education,...  UIS.X.PPP.2T3.FSGOV

   1970 [YR1970] 1971 [YR1971] 1972 [YR1972] 1973 [YR1973] 1974 [YR1974]  \
3             ..            ..            ..            ..            ..
7             ..            ..            ..            ..            ..
11            ..            ..            ..            ..            ..
15            ..            ..            ..            ..            ..
19            ..            ..            ..            ..            ..

   1975 [YR1975]  ... 2011 [YR2011] 2012 [YR2012] 2013 [YR2013] 2014 [YR2014]  \
3             ..  ...     471.78285     394.59807      579.2046     647.96967
7             ..  ...            ..            ..     214.80862            ..
11            ..  ...            ..            ..            ..            ..
15            ..  ...            ..            ..            ..            ..
```

```
19            ..  ...            ..            ..            ..            ..

   2015 [YR2015] 2016 [YR2016] 2017 [YR2017] 2018 [YR2018] 2019 [YR2019]  \
3      569.61989     695.46008      722.1059            ..            ..
7      249.26267     346.99666     291.74037            ..            ..
11            ..            ..            ..            ..            ..
15            ..            ..            ..            ..            ..
19            ..            ..            ..            ..            ..

   2020 [YR2020]
3             ..
7             ..
11            ..
15            ..
19            ..

[5 rows x 55 columns]
```

[3]:
```
df_edu_pri_ppp = df_edu_pri_ppp.melt(id_vars=["Series", "Series Code", "Country␣
↪Name","Country Code"],
       var_name="YearPriExpPPP",
       value_name="pri_exp_ppp")

df_edu_sec_ppp = df_edu_sec_ppp.melt(id_vars=["Series", "Series Code", "Country␣
↪Name","Country Code"],
       var_name="YearSecExpPPP",
       value_name="sec_exp_ppp")
```

[10]:
```
#Change year to 4 digits only.
df_edu_pri_ppp['YearPriExpPPP'] = df_edu_pri_ppp['YearPriExpPPP'].str[:4]
df_edu_sec_ppp['YearSecExpPPP'] = df_edu_sec_ppp['YearSecExpPPP'].str[:4]



#Separate % of total and % growth into 2 DFs
df_edu_pri_ppp = df_edu_pri_ppp.replace('..',np.NaN)
df_edu_sec_ppp = df_edu_sec_ppp.replace('..',np.NaN)

df_edu_pri_ppp = df_edu_pri_ppp.drop(['Series', 'Series Code'], axis = 1)
df_edu_sec_ppp = df_edu_sec_ppp.drop(['Series', 'Series Code'], axis = 1)



df_edu_pri_ppp.to_csv("../Data/Cleaned_Data/pri_exp_ppp_cleaned.csv")
df_edu_sec_ppp.to_csv("../Data/Cleaned_Data/sec_exp_ppp_cleaned.csv")
```

[ ]:

[11]:
```
df_edu_sec_ppp.head(20)
```

[11]:

	Country Name	Country Code	YearSecExpPPP	sec_exp_ppp
0	Afghanistan	AFG	1970	NaN
1	Albania	ALB	1970	NaN
2	Algeria	DZA	1970	NaN
3	American Samoa	ASM	1970	NaN
4	Andorra	AND	1970	NaN
5	Angola	AGO	1970	NaN
6	Antigua and Barbuda	ATG	1970	NaN
7	Arab World	ARB	1970	NaN
8	Argentina	ARG	1970	NaN
9	Armenia	ARM	1970	NaN
10	Aruba	ABW	1970	NaN
11	Australia	AUS	1970	NaN
12	Austria	AUT	1970	NaN
13	Azerbaijan	AZE	1970	NaN
14	Bahamas, The	BHS	1970	NaN
15	Bahrain	BHR	1970	NaN
16	Bangladesh	BGD	1970	NaN
17	Barbados	BRB	1970	NaN
18	Belarus	BLR	1970	NaN
19	Belgium	BEL	1970	NaN

[6]:
```
df_edu_sec_ppp_aus = df_edu_sec_ppp[df_edu_sec_ppp["Country Code"] == 'AUS']
#df_edu_sec_ppp_aus.head(50)
```

[7]:
```
import statistics
#print(df_edu_sec_ppp_aus.interpolate(method='linear',␣
 ↪limit_direction='forward', axis=0))
#df_edu_sec_ppp_aus.head(50)

type(df_edu_sec_ppp_aus['sec_exp_ppp'])
```

[7]:
```
pandas.core.series.Series
```

[ ]:
```

```