This notebook serves the purpose of exploring and preprocessing the “Agriculture and Rural Development GRC” dataset.

Import packages and classes

```
import pandas as pd
```

Load data

```
data_raw = pd.read_csv('../datasets/Greece - Agriculture and Rural Development/agriculture-and-rural-development_grc_raw.csv')
data_raw.head()
```

	Country Name	Country ISO3	Year	Indicator Name	Indicator Code	Value
0	#country+name	#country+code	#date+year	#indicator+name	#indicator+code	#indicator+value+num
1	Greece	GRC	2006	Agricultural machinery, tractors	AG.AGR.TRAC.NO	259613
2	Greece	GRC	2005	Agricultural machinery, tractors	AG.AGR.TRAC.NO	259766
3	Greece	GRC	2004	Agricultural machinery, tractors	AG.AGR.TRAC.NO	258476
4	Greece	GRC	2003	Agricultural machinery, tractors	AG.AGR.TRAC.NO	257737

```
data_raw.shape
```

```
(1752, 6)
```

Drop the first row of the dataset

```
data_raw = data_raw.drop(index=0)
```

We want to use the ‘Indicator Name’ variable as our features and the rest of the columns as the values. So, we use the pandas function \(pivot()\) to rotate the dataframe properly.

```
data = data_raw.pivot(index='Year', columns="Indicator Name", values="Value")
pd.set_option('display.max_columns', 5)
data.head()
```

Indicator Name	Access to electricity, rural (% of rural population)	Agricultural irrigated land (% of total agricultural land)	...	Rural population living in areas where elevation is below 5 meters (% of total population)	Surface area (sq. km)
Year					
1960	NaN	NaN	...	NaN	NaN
1961	NaN	NaN	...	NaN	131960
1962	NaN	NaN	...	NaN	131960
1963	NaN	NaN	...	NaN	131960
1964	NaN	NaN	...	NaN	131960

5 rows × 42 columns

Percentage of NaN values

```
100*data.isna().sum() / data.shape[0]
```

```
Indicator Name
Access to electricity, rural (% of rural population)
50.000000
Agricultural irrigated land (% of total agricultural land)
77.419355
Agricultural land (% of land area)
6.451613
Agricultural land (sq. km)
6.451613
Agricultural machinery, tractors
25.806452
Agricultural machinery, tractors per 100 sq. km of arable land
25.806452
Agricultural methane emissions (% of total)
37.096774
Agricultural methane emissions (thousand metric tons of CO2 equivalent)
51.612903
Agricultural nitrous oxide emissions (% of total)
37.096774
Agricultural nitrous oxide emissions (thousand metric tons of CO2 equivalent)
51.612903
Agricultural raw materials exports (% of merchandise exports)
3.225806
Agricultural raw materials imports (% of merchandise imports)
3.225806
Agriculture, forestry, and fishing, value added (% of GDP)
56.451613
Agriculture, forestry, and fishing, value added (current US$)
56.451613
Annual freshwater withdrawals, agriculture (% of total freshwater withdrawal)
83.870968
Arable land (% of land area)
6.451613
Arable land (hectares per person)
6.451613
Arable land (hectares)
6.451613
Average precipitation in depth (mm per year)
80.645161
Cereal production (metric tons)
6.451613
Cereal yield (kg per hectare)
6.451613
Crop production index (2014-2016 = 100)
4.838710
Employment in agriculture (% of total employment) (modeled ILO estimate)
53.225806
Employment in agriculture, female (% of female employment) (modeled ILO estimate)
53.225806
Employment in agriculture, male (% of male employment) (modeled ILO estimate)
53.225806
Fertilizer consumption (% of fertilizer production)
6.451613
Fertilizer consumption (kilograms per hectare of arable land)
6.451613
Food production index (2014-2016 = 100)
4.838710
Forest area (% of land area)
50.000000
Forest area (sq. km)
50.000000
Land area (sq. km)
1.612903
Land under cereal production (hectares)
6.451613
Livestock production index (2014-2016 = 100)
4.838710
Permanent cropland (% of land area)
6.451613
Rural land area (sq. km)
95.161290
Rural land area where elevation is below 5 meters (% of total land area)
95.161290
Rural land area where elevation is below 5 meters (sq. km)
95.161290
Rural population
0.000000
Rural population (% of total population)
0.000000
Rural population growth (annual %)
1.612903
Rural population living in areas where elevation is below 5 meters (% of total
population)    95.161290
Surface area (sq. km)
6.451613
dtype: float64
```

We can now save the dataframe as a cdv file using the \(to_csv()\) pandas function.

```
data = data.dropna(thresh=70*data.shape[0] /100, axis=1)
data.head()
```

Indicator Name	Agricultural land (% of land area)	Agricultural land (sq. km)	...	Rural population growth (annual %)	Surface area (sq. km)
Year					
1960	NaN	NaN	...	NaN	NaN
1961	69.1233514352211	89100	...	-0.387316115589497	131960
1962	69.0612878200155	89020	...	-1.46214309229061	131960
1963	69.9844840961986	90210	...	-1.71827774104949	131960
1964	69.7517455391777	89910	...	-1.75891977162161	131960

5 rows × 23 columns

```
data.shape
```

```
(62, 23)
```

We can now save the dataframe as a cdv file using the $to\_csv()$ pandas function.

```
data.to_csv("../datasets/Greece - Agriculture and Rural Development/agriculture-and-rural-development_grc.csv", index=False)
```