The “Greece – Agriculture and Rural Development Data” dataset sourced from the World Bank’s data portal, provides information about machinery, production, employment and cropland changes in Greece. The initial raw data contains 6 columns and 1752 samples, but we will transform the raw dataset into a format more suitable for analysis.

1.1. Import packages and classes

We will start by including the pandas library for data manipulation. The *set_option()* function is used to set the maximun number of displayable rows.

```
import pandas as pd
pd.set_option('display.max_columns', 6)
```

1.2. Load data

We can now use the pandas library to load the dataset as a pandas *DataFrame*. We also use the *info()* function to view some basic information about the dataset.

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1752 entries, 0 to 1751
Data columns (total 6 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   Country Name    1752 non-null   object
 1   Country ISO3    1752 non-null   object
 2   Year            1752 non-null   object
 3   Indicator Name  1752 non-null   object
 4   Indicator Code  1752 non-null   object
 5   Value           1752 non-null   object
dtypes: object(6)
memory usage: 82.2+ KB
```

From the summary information we can learn that the dataset has 1752 samples (rows) and 6 variables (columns). In addition, there are no *NaN* values in any of the columns and all columns are of type *‘object’*.
We can also print the first rows of the dataset using the *head()* function, just to visualize the data.

```
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

1.3. Transformations

We start by dropping the first row of the dataset, as it contains the column names.

```
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 *pivot()* pandas function to reshape the dataframe properly. We use the *‘Year’* column for the index, the *‘Indicator Name’* for the column names and the *‘Value’* column for the actual data. Again, using the *info()* and the *head()* functions, we view the summary and visualize the transformed dataset.

```
data = data_raw.pivot(index='Year', columns="Indicator Name", values="Value")
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 62 entries, 1960 to 2021
Data columns (total 42 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   Access to electricity, rural (% of rural population)
31 non-null     object
 1   Agricultural irrigated land (% of total agricultural land)
14 non-null     object
 2   Agricultural land (% of land area)
58 non-null     object
 3   Agricultural land (sq. km)
58 non-null     object
 4   Agricultural machinery, tractors
46 non-null     object
 5   Agricultural machinery, tractors per 100 sq. km of arable land
46 non-null     object
 6   Agricultural methane emissions (% of total)
39 non-null     object
 7   Agricultural methane emissions (thousand metric tons of CO2 equivalent)
30 non-null     object
 8   Agricultural nitrous oxide emissions (% of total)
39 non-null     object
 9   Agricultural nitrous oxide emissions (thousand metric tons of CO2 equivalent)
30 non-null     object
 10  Agricultural raw materials exports (% of merchandise exports)
60 non-null     object
 11  Agricultural raw materials imports (% of merchandise imports)
60 non-null     object
 12  Agriculture, forestry, and fishing, value added (% of GDP)
27 non-null     object
 13  Agriculture, forestry, and fishing, value added (current US$)
27 non-null     object
 14  Annual freshwater withdrawals, agriculture (% of total freshwater withdrawal)
10 non-null     object
 15  Arable land (% of land area)
58 non-null     object
 16  Arable land (hectares per person)
58 non-null     object
 17  Arable land (hectares)
58 non-null     object
 18  Average precipitation in depth (mm per year)
12 non-null     object
 19  Cereal production (metric tons)
58 non-null     object
 20  Cereal yield (kg per hectare)
58 non-null     object
 21  Crop production index (2014-2016 = 100)
59 non-null     object
 22  Employment in agriculture (% of total employment) (modeled ILO estimate)
29 non-null     object
 23  Employment in agriculture, female (% of female employment) (modeled ILO
estimate)           29 non-null     object
 24  Employment in agriculture, male (% of male employment) (modeled ILO estimate)
29 non-null     object
 25  Fertilizer consumption (% of fertilizer production)
58 non-null     object
 26  Fertilizer consumption (kilograms per hectare of arable land)
58 non-null     object
 27  Food production index (2014-2016 = 100)
59 non-null     object
 28  Forest area (% of land area)
31 non-null     object
 29  Forest area (sq. km)
31 non-null     object
 30  Land area (sq. km)
61 non-null     object
 31  Land under cereal production (hectares)
58 non-null     object
 32  Livestock production index (2014-2016 = 100)
59 non-null     object
 33  Permanent cropland (% of land area)
58 non-null     object
 34  Rural land area (sq. km)
3 non-null      object
 35  Rural land area where elevation is below 5 meters (% of total land area)
3 non-null      object
 36  Rural land area where elevation is below 5 meters (sq. km)
3 non-null      object
 37  Rural population
62 non-null     object
 38  Rural population (% of total population)
62 non-null     object
 39  Rural population growth (annual %)
61 non-null     object
```

```
 40  Rural population living in areas where elevation is below 5 meters (% of
total population)  3 non-null      object
 41  Surface area (sq. km)
58 non-null     object
dtypes: object(42)
memory usage: 20.8+ KB
```

```
data.head()
```

Indicator Name	Access to electricity, rural (% of rural population)	Agricultural irrigated land (% of total agricultural land)	Agricultural land (% of land area)	...	Rural population growth (annual %)	Rural population living in areas where elevation is below 5 meters (% of total population)	Surface area (sq. km)
Year							
1960	NaN	NaN	NaN	...	NaN	NaN	NaN
1961	NaN	NaN	69.1233514352211	...	-0.387316115589497	NaN	131960
1962	NaN	NaN	69.0612878200155	...	-1.46214309229061	NaN	131960
1963	NaN	NaN	69.9844840961986	...	-1.71827774104949	NaN	131960
1964	NaN	NaN	69.7517455391777	...	-1.75891977162161	NaN	131960

5 rows × 42 columns

We calculate the percentage of *NaN* values in each column by using the *isna()* and *sum()* functions.

```
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 set a threshold to drop columns that have a high *NaN* percentage. We use the *dropna()* function and set the threshold to 50%, so that columns with more than 50% of *NaN* values will be dropped.

1.4. Plots

1.5. Store Dataset

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)
```