In [59]:
```
import numpy as np
import pandas as pd

df_yield = pd.read_csv('yield1.csv')
df_yield.shape
```

Out[59]:
```
(108971, 7)
```

In [60]:
```
df_yield.head()
```

Out[60]:

| | Domain | Area | Element | Item | Year | Unit | Value |
|---|---|---|---|---|---|---|---|
| **0** | Crops | Afghanistan | Yield | Apples | 1961 | hg/ha | 68018.0 |
| **1** | Crops | Afghanistan | Yield | Apples | 1962 | hg/ha | 68018.0 |
| **2** | Crops | Afghanistan | Yield | Apples | 1963 | hg/ha | 68018.0 |
| **3** | Crops | Afghanistan | Yield | Apples | 1964 | hg/ha | 78298.0 |
| **4** | Crops | Afghanistan | Yield | Apples | 1965 | hg/ha | 82258.0 |

In [61]:
```
df_yield = df_yield.drop(['Element','Domain','Unit'], axis=1)
df_yield.head()
```

Out[61]:

| | Area | Item | Year | Value |
|---|---|---|---|---|
| **0** | Afghanistan | Apples | 1961 | 68018.0 |
| **1** | Afghanistan | Apples | 1962 | 68018.0 |
| **2** | Afghanistan | Apples | 1963 | 68018.0 |
| **3** | Afghanistan | Apples | 1964 | 78298.0 |
| **4** | Afghanistan | Apples | 1965 | 82258.0 |

In [62]:
```
df_yield = df_yield.rename(columns={"Value": "hg/ha_yield"})
df_yield.head()
```

Out[62]:

| | Area | Item | Year | hg/ha_yield |
|---|---|---|---|---|
| **0** | Afghanistan | Apples | 1961 | 68018.0 |
| **1** | Afghanistan | Apples | 1962 | 68018.0 |
| **2** | Afghanistan | Apples | 1963 | 68018.0 |
| **3** | Afghanistan | Apples | 1964 | 78298.0 |
| **4** | Afghanistan | Apples | 1965 | 82258.0 |

In [63]:
```
df_yield.isnull().sum()
```

Out[63]:
```
Area           0
Item           0
Year           0
hg/ha_yield    6
dtype: int64
```

In [64]:
```
df_yield.dropna(axis=0, how='any', inplace=True)
df_yield.isnull().sum()
```

Out[64]:
```
Area           0
Item           0
Year           0
hg/ha_yield    0
dtype: int64
```

In [65]:
```
df_yield.shape
```

Out[65]:
```
(108965, 4)
```

In [66]:
```
df_yield.describe()
```

Out[66]:

| | Year | hg/ha_yield |
|---|---|---|
| **count** | 108965.000000 | 1.089650e+05 |
| **mean** | 1991.546708 | 8.887500e+04 |
| **std** | 16.613099 | 1.570503e+05 |
| **min** | 1961.000000 | 0.000000e+00 |
| **25%** | 1977.000000 | 1.163300e+04 |
| **50%** | 1993.000000 | 3.333300e+04 |
| **75%** | 2006.000000 | 9.240900e+04 |
| **max** | 2018.000000 | 2.981628e+06 |

In [67]:
```
df_yield.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 108965 entries, 0 to 108970
Data columns (total 4 columns):
 #   Column       Non-Null Count   Dtype  
---  ------       --------------   -----  
 0   Area         108965 non-null  object 
 1   Item         108965 non-null  object 
 2   Year         108965 non-null  int64  
 3   hg/ha_yield  108965 non-null  float64
dtypes: float64(1), int64(1), object(2)
memory usage: 4.2+ MB
```

# area harvested

In [70]:
```
df_area = pd.read_csv('area.csv')
df_area.head()
```

Out[70]:

| | Domain | Area | Element | Item | Year | Unit | Value |
|---|---|---|---|---|---|---|---|
| **0** | Crops | Afghanistan | Area harvested | Apples | 1961 | ha | 2220.0 |
| **1** | Crops | Afghanistan | Area harvested | Apples | 1962 | ha | 2220.0 |
| **2** | Crops | Afghanistan | Area harvested | Apples | 1963 | ha | 2220.0 |
| **3** | Crops | Afghanistan | Area harvested | Apples | 1964 | ha | 2350.0 |
| **4** | Crops | Afghanistan | Area harvested | Apples | 1965 | ha | 2480.0 |

In [71]:
```
df_area = df_area.drop(['Element','Domain','Unit'], axis=1)
df_area.head()
```

Out[71]:

| | Area | Item | Year | Value |
|---|---|---|---|---|
| **0** | Afghanistan | Apples | 1961 | 2220.0 |
| **1** | Afghanistan | Apples | 1962 | 2220.0 |
| **2** | Afghanistan | Apples | 1963 | 2220.0 |
| **3** | Afghanistan | Apples | 1964 | 2350.0 |
| **4** | Afghanistan | Apples | 1965 | 2480.0 |

In [72]:
```
df_area.isnull().sum()
```

Out[72]:
```
Area         0
Item         0
Year         0
Value    11325
dtype: int64
```

In [73]:
```
df_area.dropna(axis=0, how='any', inplace=True)
df_area.isnull().sum()
```

Out[73]:
```
Area     0
Item     0
Year     0
Value    0
dtype: int64
```

In [74]:
```
df_area.shape
```

Out[74]:
```
(109467, 4)
```

In [75]:
```
df_area.describe()
```

Out[75]:

| | Year | Value |
|---|---|---|
| **count** | 109467.000000 | 1.094670e+05 |
| **mean** | 1991.582541 | 4.001220e+05 |
| **std** | 16.599720 | 2.528810e+06 |
| **min** | 1961.000000 | 0.000000e+00 |
| **25%** | 1978.000000 | 1.200000e+03 |
| **50%** | 1993.000000 | 9.200000e+03 |
| **75%** | 2006.000000 | 6.590550e+04 |
| **max** | 2018.000000 | 7.020501e+07 |

In [76]:
```
df_area.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 109467 entries, 0 to 120791
Data columns (total 4 columns):
 #   Column  Non-Null Count   Dtype  
---  ------  --------------   -----  
 0   Area    109467 non-null  object 
 1   Item    109467 non-null  object 
 2   Year    109467 non-null  int64  
 3   Value   109467 non-null  float64
dtypes: float64(1), int64(1), object(2)
memory usage: 4.2+ MB
```

In [77]:
```
df_area = df_area.rename(columns={"Value": "ha_area"})
df_area.head()
```

Out[77]:

| | Area | Item | Year | ha_area |
|---|---|---|---|---|
| **0** | Afghanistan | Apples | 1961 | 2220.0 |
| **1** | Afghanistan | Apples | 1962 | 2220.0 |
| **2** | Afghanistan | Apples | 1963 | 2220.0 |
| **3** | Afghanistan | Apples | 1964 | 2350.0 |
| **4** | Afghanistan | Apples | 1965 | 2480.0 |

In [78]:
```
yield_df = pd.merge(df_yield, df_area, on=['Year','Area','Item'])
yield_df.head()
```

Out[78]:

| | Area | Item | Year | hg/ha_yield | ha_area |
|---|---|---|---|---|---|
| **0** | Afghanistan | Apples | 1961 | 68018.0 | 2220.0 |
| **1** | Afghanistan | Apples | 1962 | 68018.0 | 2220.0 |
| **2** | Afghanistan | Apples | 1963 | 68018.0 | 2220.0 |
| **3** | Afghanistan | Apples | 1964 | 78298.0 | 2350.0 |
| **4** | Afghanistan | Apples | 1965 | 82258.0 | 2480.0 |

In [79]:
```
yield_df.shape
```

Out[79]:
```
(108965, 5)
```

# rainfall data

In [80]:
```
df_rain = pd.read_csv('rainfall.csv')
df_rain.head()
```

Out[80]:

| | Area | Year | average_rain_fall_mm_per_year |
|---|---|---|---|
| **0** | Afghanistan | 1985 | 327.0 |
| **1** | Afghanistan | 1986 | 327.0 |
| **2** | Afghanistan | 1987 | 327.0 |
| **3** | Afghanistan | 1989 | 327.0 |
| **4** | Afghanistan | 1990 | 327.0 |

In [81]:
```
df_rain.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6727 entries, 0 to 6726
Data columns (total 3 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   Area                            6727 non-null   object 
 1   Year                            6727 non-null   int64  
 2   average_rain_fall_mm_per_year  5947 non-null   float64
dtypes: float64(1), int64(1), object(1)
memory usage: 157.8+ KB
```

In [82]:
```
df_rain.head()
```

Out[82]:

| | Area | Year | average_rain_fall_mm_per_year |
|---|---|---|---|
| **0** | Afghanistan | 1985 | 327.0 |
| **1** | Afghanistan | 1986 | 327.0 |
| **2** | Afghanistan | 1987 | 327.0 |
| **3** | Afghanistan | 1989 | 327.0 |
| **4** | Afghanistan | 1990 | 327.0 |

In [83]:
```
df_rain.isnull().sum()
```

Out[83]:
```
Area                              0
Year                              0
average_rain_fall_mm_per_year    780
dtype: int64
```

In [84]:
```
print(df_rain.shape)
df_rain.dropna(axis=0, how='any', inplace=True)
df_rain.isnull().sum()
```

```
(6727, 3)
```

Out[84]:
```
Area                              0
Year                              0
average_rain_fall_mm_per_year    0
dtype: int64
```

In [85]:
```
df_rain.shape
```

Out[85]:
```
(5947, 3)
```

In [86]:
```
yield_df = pd.merge(yield_df, df_rain, on=['Year','Area'])
yield_df.head()
```

Out[86]:

| | Area | Item | Year | hg/ha_yield | ha_area | average_rain_fall_mm_per_year |
|---|---|---|---|---|---|---|
| **0** | Afghanistan | Apples | 1985 | 71805.0 | 2660.0 | 327.0 |
| **1** | Afghanistan | Grapes | 1985 | 72340.0 | 56400.0 | 327.0 |
| **2** | Afghanistan | Maize | 1985 | 16652.0 | 346500.0 | 327.0 |
| **3** | Afghanistan | Millet | 1985 | 8710.0 | 31000.0 | 327.0 |
| **4** | Afghanistan | Oranges | 1985 | 75294.0 | 1700.0 | 327.0 |

In [87]:
```
yield_df.shape
```

Out[87]:
```
(50521, 6)
```

In [88]:
```
yield_df.describe()
```

Out[88]:

| | Year | hg/ha_yield | ha_area | average_rain_fall_mm_per_year |
|---|---|---|---|---|
| **count** | 50521.000000 | 5.052100e+04 | 5.052100e+04 | 50521.000000 |
| **mean** | 2001.947032 | 9.414374e+04 | 3.429823e+05 | 1208.583698 |
| **std** | 9.401671 | 1.615778e+05 | 2.056346e+06 | 814.602694 |
| **min** | 1985.000000 | 5.600000e+01 | 0.000000e+00 | 51.000000 |
| **25%** | 1994.000000 | 1.338700e+04 | 1.257000e+03 | 591.000000 |
| **50%** | 2002.000000 | 3.707600e+04 | 9.100000e+03 | 1083.000000 |
| **75%** | 2010.000000 | 9.853600e+04 | 6.273600e+04 | 1738.000000 |
| **max** | 2017.000000 | 2.942959e+06 | 4.553740e+07 | 3240.000000 |

# pesticides

In [89]:
```
df_pes = pd.read_csv('pests.csv')
df_pes.head()
```

Out[89]:

| | Domain | Area | Element | Item | Year | Unit | Value |
|---|---|---|---|---|---|---|---|
| **0** | Pesticides indicators | Albania | Use per area of cropland | Pesticides (total) | 1990 | kg/ha | 0.17 |
| **1** | Pesticides indicators | Albania | Use per area of cropland | Pesticides (total) | 1991 | kg/ha | 0.17 |
| **2** | Pesticides indicators | Albania | Use per area of cropland | Pesticides (total) | 1992 | kg/ha | 0.17 |
| **3** | Pesticides indicators | Albania | Use per area of cropland | Pesticides (total) | 1993 | kg/ha | 0.17 |
| **4** | Pesticides indicators | Albania | Use per area of cropland | Pesticides (total) | 1994 | kg/ha | 0.29 |

In [90]:
```
df_pes = df_pes.rename(columns={"Value": "pestc_kg/ha"})
df_pes = df_pes.drop(['Element','Domain','Unit','Item'], axis=1)
df_pes.head()
```

Out[90]:

| | Area | Year | pestc_kg/ha |
|---|---|---|---|
| **0** | Albania | 1990 | 0.17 |
| **1** | Albania | 1991 | 0.17 |
| **2** | Albania | 1992 | 0.17 |
| **3** | Albania | 1993 | 0.17 |
| **4** | Albania | 1994 | 0.29 |

In [91]:
```
print(df_pes.shape)
df_pes.isnull().sum()
```

```
(4287, 3)
```

Out[91]:
```
Area           0
Year           0
pestc_kg/ha    0
dtype: int64
```

In [92]:
```
yield_df = pd.merge(yield_df, df_pes, on=['Year','Area'])
yield_df.shape
```

Out[92]:
```
(37615, 7)
```

In [93]:
```
yield_df.head()
```

Out[93]:

| | Area | Item | Year | hg/ha_yield | ha_area | average_rain_fall_mm_per_year | pestc_kg/ha |
|---|---|---|---|---|---|---|---|
| **0** | Albania | Apples | 1990 | 50385.0 | 2600.0 | 1485.0 | 0.17 |
| **1** | Albania | Grapes | 1990 | 64518.0 | 14058.0 | 1485.0 | 0.17 |
| **2** | Albania | Maize | 1990 | 36613.0 | 62000.0 | 1485.0 | 0.17 |
| **3** | Albania | Oranges | 1990 | 60000.0 | 1700.0 | 1485.0 | 0.17 |
| **4** | Albania | Rice, paddy | 1990 | 23333.0 | 3000.0 | 1485.0 | 0.17 |

# temperature

In [94]:
```
avg_temp=  pd.read_csv('temp.csv')
avg_temp.head()
```

Out[94]:

| | year | country | avg_temp |
|---|---|---|---|
| **0** | 1849 | Côte D'Ivoire | 25.58 |
| **1** | 1850 | Côte D'Ivoire | 25.52 |
| **2** | 1851 | Côte D'Ivoire | 25.67 |
| **3** | 1852 | Côte D'Ivoire | NaN |
| **4** | 1853 | Côte D'Ivoire | NaN |

In [95]:
```
avg_temp.shape
```

Out[95]:
```
(71311, 3)
```

In [96]:
```
avg_temp.isnull().sum()
```

Out[96]:
```
year           0
country        0
avg_temp    2547
dtype: int64
```

In [97]:
```
avg_temp.dropna(axis=0, how='any', inplace=True)
avg_temp = avg_temp.rename(columns={"year": "Year", "country":'Area'})
avg_temp.isnull().sum()
```

Out[97]:
```
Year        0
Area        0
avg_temp    0
dtype: int64
```

In [98]:
```
avg_temp.shape
```

Out[98]:
```
(68764, 3)
```

In [99]:
```
yield_df = pd.merge(yield_df,avg_temp, on=['Area','Year'])
yield_df.head()
```

Out[99]:

| | Area | Item | Year | hg/ha_yield | ha_area | average_rain_fall_mm_per_year | pestc_kg/ha | avg_t |
|---|---|---|---|---|---|---|---|---|
| 0 | Albania | Apples | 1990 | 50385.0 | 2600.0 | 1485.0 | 0.17 | |
| 1 | Albania | Grapes | 1990 | 64518.0 | 14058.0 | 1485.0 | 0.17 | |
| 2 | Albania | Maize | 1990 | 36613.0 | 62000.0 | 1485.0 | 0.17 | |
| 3 | Albania | Oranges | 1990 | 60000.0 | 1700.0 | 1485.0 | 0.17 | |
| 4 | Albania | Rice, paddy | 1990 | 23333.0 | 3000.0 | 1485.0 | 0.17 | |

In [100]:
```
yield_df.shape
```

Out[100]:
```
(81113, 8)
```

# fertilizers

In [101]:
```
# nitrogen fertilizer
fert_n = pd.read_csv('fert_N.csv')
fert_n.head()
```

Out[101]:

| | Domain | Area | Element | Item | Year | Unit | Value |
|---|---|---|---|---|---|---|---|
| 0 | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient nitrogen N (total) | 2002 | kg/ha | 3.16 |
| 1 | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient nitrogen N (total) | 2003 | kg/ha | 2.58 |
| 2 | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient nitrogen N (total) | 2004 | kg/ha | 2.82 |
| 3 | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient nitrogen N (total) | 2005 | kg/ha | 2.59 |
| 4 | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient nitrogen N (total) | 2006 | kg/ha | 2.59 |

In [102]:
```
fert_n = fert_n.drop(['Element','Domain','Unit','Item'], axis=1)
fert_n.head()
```

Out[102]:

| | Area | Year | Value |
|---|---|---|---|
| **0** | Afghanistan | 2002 | 3.16 |
| **1** | Afghanistan | 2003 | 2.58 |
| **2** | Afghanistan | 2004 | 2.82 |
| **3** | Afghanistan | 2005 | 2.59 |
| **4** | Afghanistan | 2006 | 2.59 |

In [103]:
```
fert_n = fert_n.rename(columns={"Value" : "fertN_kg/ha"})
```

In [104]:
```
fert_n.isnull().sum()
```

Out[104]:
```
Area           0
Year           0
fertN_kg/ha    0
dtype: int64
```

In [105]:
```
yield_df = pd.merge(yield_df, fert_n, on=['Year','Area'])
yield_df.head()
```

Out[105]:

| | Area | Item | Year | hg/ha_yield | ha_area | average_rain_fall_mm_per_year | pestc_kg/ha | avg_t |
|---|---|---|---|---|---|---|---|---|
| **0** | Albania | Apples | 2002 | 70222.0 | 2250.0 | 1485.0 | 0.47 | |
| **1** | Albania | Beans, green | 2002 | 78261.0 | 690.0 | 1485.0 | 0.47 | |
| **2** | Albania | Grapes | 2002 | 159746.0 | 5202.0 | 1485.0 | 0.47 | |
| **3** | Albania | Maize | 2002 | 39460.0 | 50000.0 | 1485.0 | 0.47 | |
| **4** | Albania | Oranges | 2002 | 68000.0 | 500.0 | 1485.0 | 0.47 | |

In [106]:
```
# phosphorous fertilizer
fert_p = pd.read_csv('fert_P.csv')
fert_p.head()
```

Out[106]:

| | Domain | Area | Element | Item | Year | Unit | Value |
|---|---|---|---|---|---|---|---|
| **0** | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient phosphate P2O5 (total) | 2002 | kg/ha | 0.00 |
| **1** | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient phosphate P2O5 (total) | 2003 | kg/ha | 0.84 |
| **2** | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient phosphate P2O5 (total) | 2004 | kg/ha | 1.36 |
| **3** | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient phosphate P2O5 (total) | 2005 | kg/ha | 1.16 |
| **4** | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient phosphate P2O5 (total) | 2006 | kg/ha | 0.56 |

In [107]:
```
fert_p = fert_p.drop(['Element','Domain','Unit','Item'], axis=1)
fert_p.head()
```

Out[107]:

| | Area | Year | Value |
|---|---|---|---|
| **0** | Afghanistan | 2002 | 0.00 |
| **1** | Afghanistan | 2003 | 0.84 |
| **2** | Afghanistan | 2004 | 1.36 |
| **3** | Afghanistan | 2005 | 1.16 |
| **4** | Afghanistan | 2006 | 0.56 |

In [108]:
```
fert_p = fert_p.rename(columns={"Value" : "fertP_kg/ha"})
```

In [109]:
```
fert_p.isnull().sum()
```

Out[109]:
```
Area           0
Year           0
fertP_kg/ha    0
dtype: int64
```

In [110]:
```
yield_df = pd.merge(yield_df, fert_p, on=['Year','Area'])
yield_df.head()
```

Out[110]:

| | Area | Item | Year | hg/ha_yield | ha_area | average_rain_fall_mm_per_year | pestc_kg/ha | avg_t |
|---|---|---|---|---|---|---|---|---|
| **0** | Albania | Apples | 2002 | 70222.0 | 2250.0 | 1485.0 | 0.47 | |
| **1** | Albania | Beans, green | 2002 | 78261.0 | 690.0 | 1485.0 | 0.47 | |
| **2** | Albania | Grapes | 2002 | 159746.0 | 5202.0 | 1485.0 | 0.47 | |
| **3** | Albania | Maize | 2002 | 39460.0 | 50000.0 | 1485.0 | 0.47 | |
| **4** | Albania | Oranges | 2002 | 68000.0 | 500.0 | 1485.0 | 0.47 | |

In [111]:
```
# potassium fertilizer
fert_k = pd.read_csv('fert_K.csv')
fert_k.head()
```

Out[111]:

| | Domain | Area | Element | Item | Year | Unit | Value |
|---|---|---|---|---|---|---|---|
| **0** | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient potash K2O (total) | 2002 | kg/ha | 0.00 |
| **1** | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient potash K2O (total) | 2003 | kg/ha | 0.00 |
| **2** | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient potash K2O (total) | 2004 | kg/ha | 0.00 |
| **3** | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient potash K2O (total) | 2005 | kg/ha | 0.01 |
| **4** | Fertilizers indicators | Afghanistan | Use per area of cropland | Nutrient potash K2O (total) | 2006 | kg/ha | 0.00 |

In [112]:
```
fert_k = fert_k.drop(['Element','Domain','Unit','Item'], axis=1)
fert_k.head()
```

Out[112]:

| | Area | Year | Value |
|---|---|---|---|
| **0** | Afghanistan | 2002 | 0.00 |
| **1** | Afghanistan | 2003 | 0.00 |
| **2** | Afghanistan | 2004 | 0.00 |
| **3** | Afghanistan | 2005 | 0.01 |
| **4** | Afghanistan | 2006 | 0.00 |

In [113]:
```
fert_k = fert_k.rename(columns={"Value" : "fertK_kg/ha"})
```

In [114]:
```
fert_k.isnull().sum()
```

Out[114]:
```
Area           0
Year           0
fertK_kg/ha    0
dtype: int64
```

In [115]:
```
yield_df = pd.merge(yield_df, fert_k, on=['Year','Area'])
yield_df.head()
```

Out[115]:

| | Area | Item | Year | hg/ha_yield | ha_area | average_rain_fall_mm_per_year | pestc_kg/ha | avg_t |
|---|---|---|---|---|---|---|---|---|
| **0** | Albania | Apples | 2002 | 70222.0 | 2250.0 | 1485.0 | 0.47 | 1 |
| **1** | Albania | Beans, green | 2002 | 78261.0 | 690.0 | 1485.0 | 0.47 | 1 |
| **2** | Albania | Grapes | 2002 | 159746.0 | 5202.0 | 1485.0 | 0.47 | 1 |
| **3** | Albania | Maize | 2002 | 39460.0 | 50000.0 | 1485.0 | 0.47 | 1 |
| **4** | Albania | Oranges | 2002 | 68000.0 | 500.0 | 1485.0 | 0.47 | 1 |

In [116]:
```
def func(a):
    s=a.split(',')
    return s[0]

a=list(map(func, list(yield_df['Item'])))

yield_df['Item']=a
```

In [117]:
```
print(yield_df.shape)
yield_df.head()
```

```
(38320, 11)
```

Out[117]:

| | Area | Item | Year | hg/ha_yield | ha_area | average_rain_fall_mm_per_year | pestc_kg/ha | avg_t |
|---|---|---|---|---|---|---|---|---|
| **0** | Albania | Apples | 2002 | 70222.0 | 2250.0 | 1485.0 | 0.47 | |
| **1** | Albania | Beans | 2002 | 78261.0 | 690.0 | 1485.0 | 0.47 | |
| **2** | Albania | Grapes | 2002 | 159746.0 | 5202.0 | 1485.0 | 0.47 | |
| **3** | Albania | Maize | 2002 | 39460.0 | 50000.0 | 1485.0 | 0.47 | |
| **4** | Albania | Oranges | 2002 | 68000.0 | 500.0 | 1485.0 | 0.47 | |

In [118]:
```
yield_df.isnull().sum()
```

Out[118]:

```
Area                              0
Item                              0
Year                              0
hg/ha_yield                       0
ha_area                           0
average_rain_fall_mm_per_year     0
pestc_kg/ha                       0
avg_temp                          0
fertN_kg/ha                       0
fertP_kg/ha                       0
fertK_kg/ha                       0
dtype: int64
```

In [119]:
```
yield_df.describe()
```

Out[119]:

| | Year | hg/ha_yield | ha_area | average_rain_fall_mm_per_year | pestc_kg/ha | |
|---|---|---|---|---|---|---|
| **count** | 38320.000000 | 3.832000e+04 | 3.832000e+04 | 38320.000000 | 38320.000000 | 383 |
| **mean** | 2007.943267 | 1.130228e+05 | 2.098866e+06 | 998.548121 | 4.657181 | |
| **std** | 3.317127 | 1.708848e+05 | 6.343620e+06 | 613.610247 | 5.365177 | |
| **min** | 2002.000000 | 1.620000e+02 | 0.000000e+00 | 51.000000 | 0.000000 | |
| **25%** | 2005.000000 | 1.884775e+04 | 1.052900e+04 | 636.000000 | 0.250000 | |
| **50%** | 2008.000000 | 4.796800e+04 | 9.938300e+04 | 686.000000 | 1.950000 | |
| **75%** | 2011.000000 | 1.275590e+05 | 7.882170e+05 | 1180.000000 | 10.450000 | |
| **max** | 2013.000000 | 2.799434e+06 | 4.553740e+07 | 3142.000000 | 14.860000 | |

In [120]:
```
yield_df.shape
```

Out[120]:

```
(38320, 11)
```

In [122]:
```
yield_df.to_csv('yield_df.csv')
```

In [ ]:
```

```