

# Pesticide Usage Dataset Analysis Overview

This analysis provides insights into a dataset containing information about pesticide usage across different domains, areas, elements, items, years, units, and values. The objective is to identify trends, patterns, and outliers in pesticide usage to better understand how and where pesticides are being applied.

## Dataset Columns

- **Domain:** The category of usage (e.g., agricultural, industrial).
- **Area:** The geographical region where the pesticide is used.
- **Element:** The chemical component of the pesticide.
- **Item:** The specific product or substance.
- **Year:** The year of data collection.
- **Unit:** The measurement unit for the pesticide value.
- **Value:** The quantity of pesticide usage.

## Key Questions and Insights

### 1. Which areas have the highest pesticide usage?

By aggregating pesticide usage values by area, we identify the top regions where pesticides are most heavily used. This helps in focusing regulatory and environmental efforts.

### 2. How has pesticide usage changed over the years?

Analyzing the yearly pesticide usage trends reveals whether there is an increase or decrease in pesticide application over time, indicating the effectiveness of policies and changes in agricultural practices.

### 3. What are the top elements contributing to pesticide usage?

Identifying the most used chemical elements in pesticides can help in assessing potential environmental and health risks associated with sustainable practices.

### 5. How does pesticide usage vary between different domains?

A domain-wise breakdown of pesticide usage helps in understanding how different sectors contribute to overall pesticide application, facilitating different datasets and studies.

### 7. Are there any noticeable trends or patterns in pesticide usage across different areas and items?

A heatmap visualization of pesticide usage across various areas and items can reveal hidden patterns and correlations that are not immediately obvious from raw data.

## Additional Insights

### 8. Which years have seen the highest increase or decrease in pesticide usage?

Year-over-year percentage changes highlight periods of significant increases or decreases, providing context for external factors influencing pesticide usage.

### 9. What is the average pesticide usage per area?

Calculating the average pesticide usage per area helps in understanding the typical levels of pesticide application, useful for benchmarking and identifying outliers.

### 10. Is there any seasonality in pesticide usage over the years?

Seasonal trends in pesticide usage can inform the timing of pesticide application and dependency of pesticide usage within each domain.

### 14. Which combination of area and element has the highest pesticide usage?

Finding the top area-element combinations reveals the most critical intersections that require attention for regulatory measures and

environmental protection.

## Visualization Theme

All visualizations use a green color theme to maintain consistency and emphasize the environmental aspect of pesticide usage.

## Conclusion

This comprehensive analysis of the pesticide usage dataset provides valuable insights into various dimensions of pesticide application. By understanding these patterns, stakeholders can make informed decisions to promote sustainable practices and mitigate potential risks associated with pesticide use.

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from itertools import combinations
from collections import Counter

In [3]: df = pd.read_csv("pesticides.csv")

In [5]: df.columns

Out[5]: Index(['Domain', 'Area', 'Element', 'Item', 'Year', 'Unit', 'Value'], dtype='object')

In [7]: df.shape

Out[7]: (4349, 7)

In [11]: df.head()

Out[11]:
```

	Domain	Area	Element	Item	Year	Unit	Value
0	Pesticides Use	Albania	Use	Pesticides (total)	1990	tonnes of active ingredients	121.0
1	Pesticides Use	Albania	Use	Pesticides (total)	1991	tonnes of active ingredients	121.0
2	Pesticides Use	Albania	Use	Pesticides (total)	1992	tonnes of active ingredients	121.0
3	Pesticides Use	Albania	Use	Pesticides (total)	1993	tonnes of active ingredients	121.0
4	Pesticides Use	Albania	Use	Pesticides (total)	1994	tonnes of active ingredients	201.0

```
In [13]: df . isnull(). sum()

Out[13]: Domain      0
Area      0
Element    0
Item      0
Year      0
Unit      0
Value     0
dtype: int64

In [15]: print(df.dtypes)

Domain      object
Area        object
Element     object
Item        object
Year        int64
Unit        object
Value       float64
dtype: object

In [17]: print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4349 entries, 0 to 4348
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Domain      4349 non-null   object
1   Area        4349 non-null   object
2   Element     4349 non-null   object
3   Item        4349 non-null   object
4   Year        4349 non-null   int64
5   Unit        4349 non-null   object
6   Value       4349 non-null   float64
dtypes: float64(1), int64(1), object(5)
memory usage: 238.0+ KB
None
```

which area has the highest pesticide usage?

```
In [40]: # Define the green color palette
green_color = '#2ca02c'

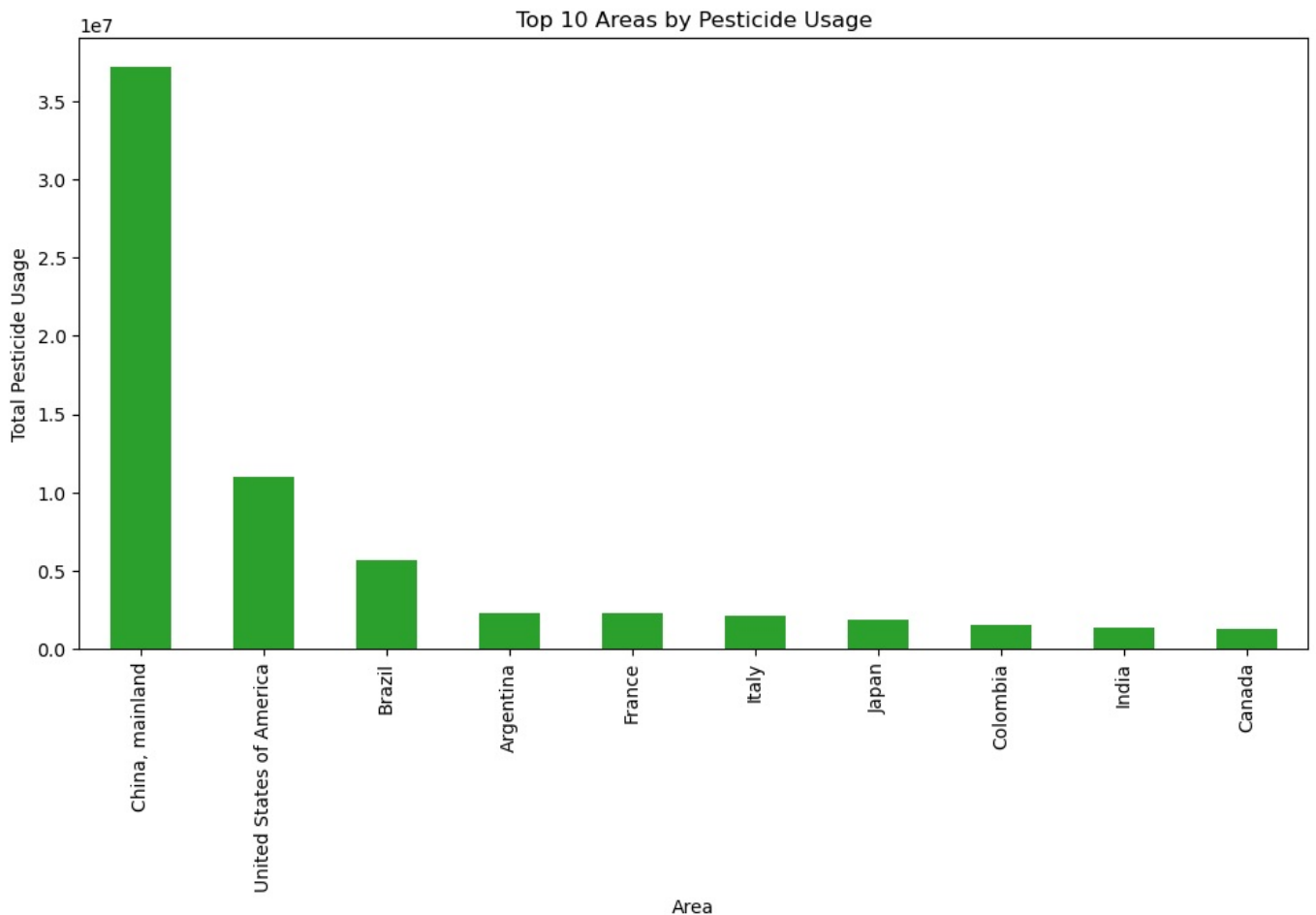
# 1. Which areas have the highest pesticide usage?
area_usage = df.groupby('Area')['Value'].sum().sort_values(ascending=False)
print("Top areas by pesticide usage:\n", area_usage.head(10))

# Plotting the top 10 areas
plt.figure(figsize=(12, 6))
area_usage.head(10).plot(kind='bar', color=green_color)
plt.title('Top 10 Areas by Pesticide Usage')
plt.xlabel('Area')
plt.ylabel('Total Pesticide Usage')
plt.show()
```

Top areas by pesticide usage:

Area	Value
China, mainland	37194836.00
United States of America	10978289.10
Brazil	5671552.00
Argentina	2250064.53
France	2248406.72
Italy	2113106.60
Japan	1843156.79
Colombia	1481230.51
India	1322613.00
Canada	1253897.98

Name: Value, dtype: float64



## How has pesticide usage changed over the years?

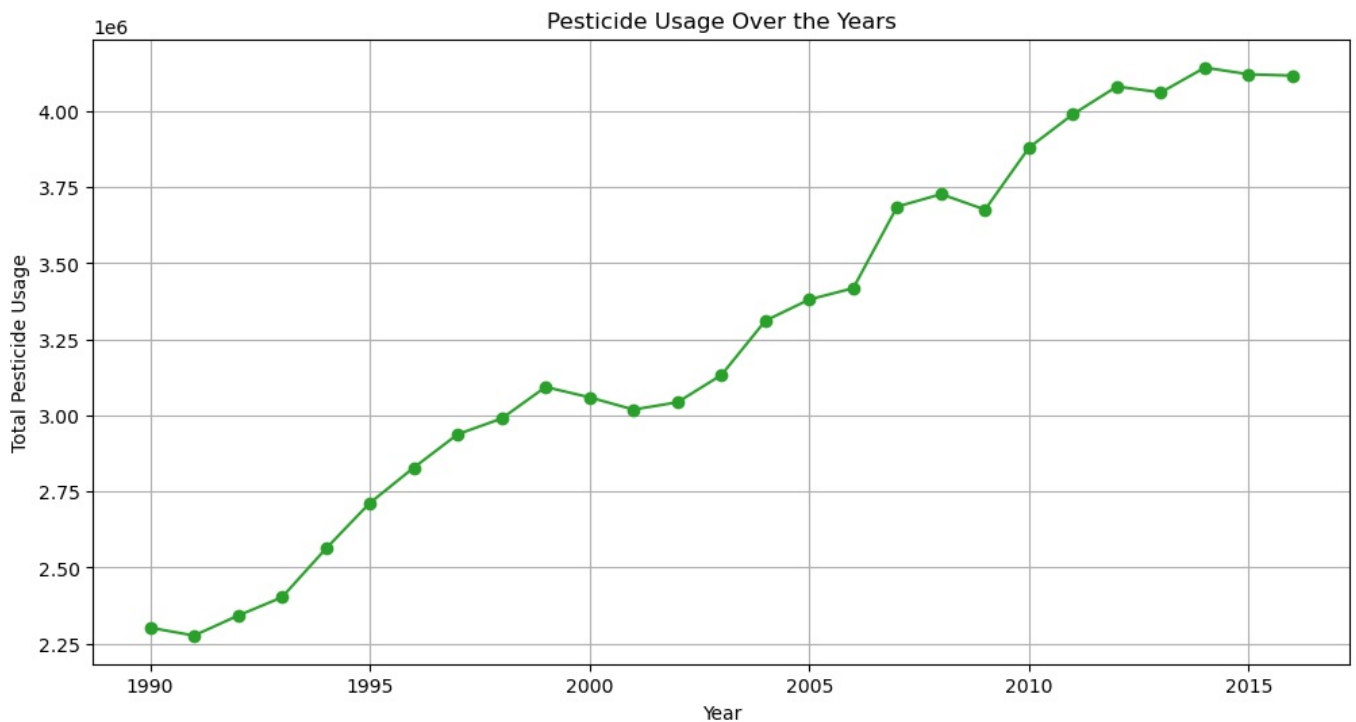
```
In [42]: # 2. How has pesticide usage changed over the years?
yearly_usage = df.groupby('Year')['Value'].sum()
print("Pesticide usage over the years:\n", yearly_usage)

# Plotting the usage over years
plt.figure(figsize=(12, 6))
yearly_usage.plot(kind='line', marker='o', color='green_color')
plt.title('Pesticide Usage Over the Years')
plt.xlabel('Year')
plt.ylabel('Total Pesticide Usage')
plt.grid(True)
plt.show()
```

Pesticide usage over the years:

Year	Value
1990	2302488.17
1991	2276735.88
1992	2342269.46
1993	2402384.07
1994	2562417.10
1995	2712360.49
1996	2828473.75
1997	2938378.10
1998	2989981.91
1999	3093701.63
2000	3059522.02
2001	3019373.93
2002	3043312.03
2003	3132912.52
2004	3311319.29
2005	3381107.30
2006	3418241.96
2007	3686339.00
2008	3727261.76
2009	3676413.63
2010	3880328.14
2011	3990014.91
2012	4081081.95
2013	4061557.91
2014	4143202.61
2015	4121220.10
2016	4116832.41

Name: Value, dtype: float64



## What are the top elements contributing to pesticide usage?

```
In [44]: # 3. What are the top elements contributing to pesticide usage?
element_usage = df.groupby('Element')['Value'].sum().sort_values(ascending=False)
print("Top elements by pesticide usage:\n", element_usage.head(10))

# Plotting the top 10 elements
```

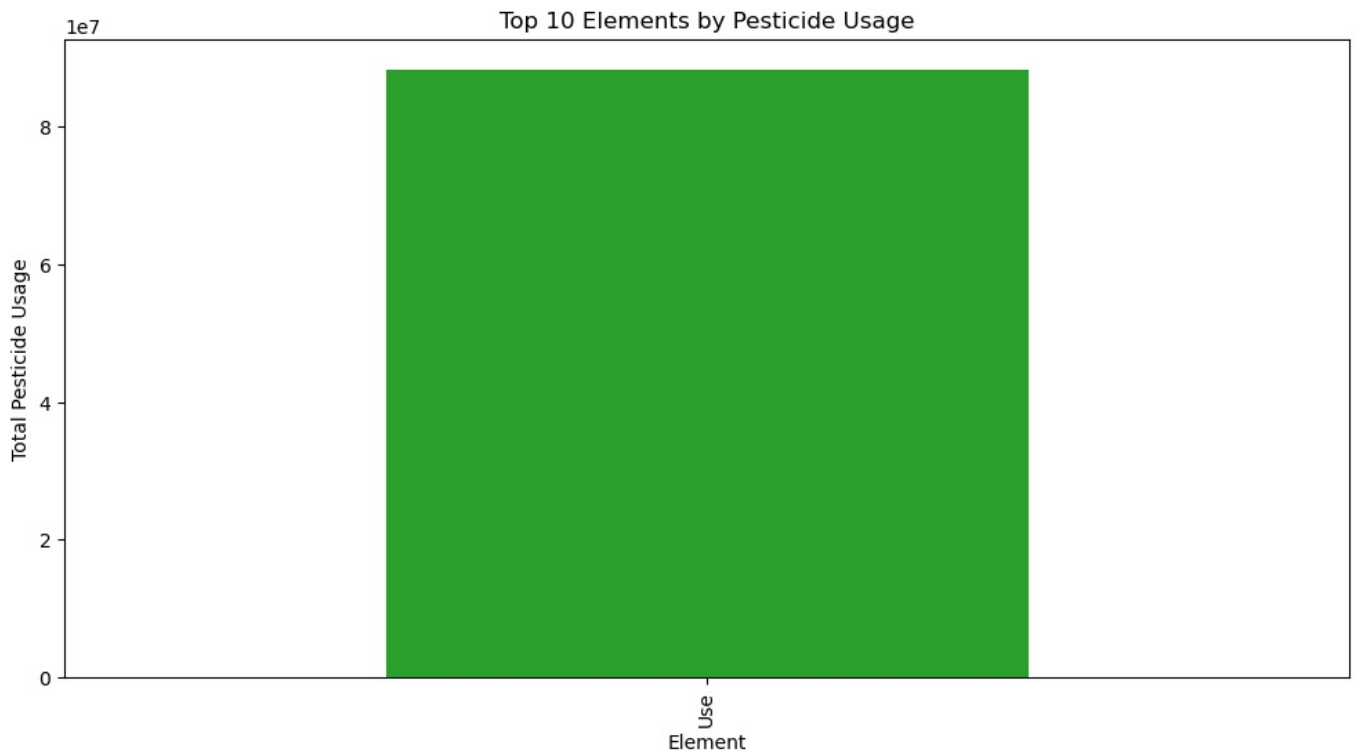
```
plt.figure(figsize=(12, 6))
element_usage.head(10).plot(kind='bar', color=green_color)
plt.title('Top 10 Elements by Pesticide Usage')
plt.xlabel('Element')
plt.ylabel('Total Pesticide Usage')
plt.show()
```

Top elements by pesticide usage:

Element

Use 88299232.03

Name: Value, dtype: float64



# 7. Are there any noticeable trends or patterns in pesticide usage across different areas and items?

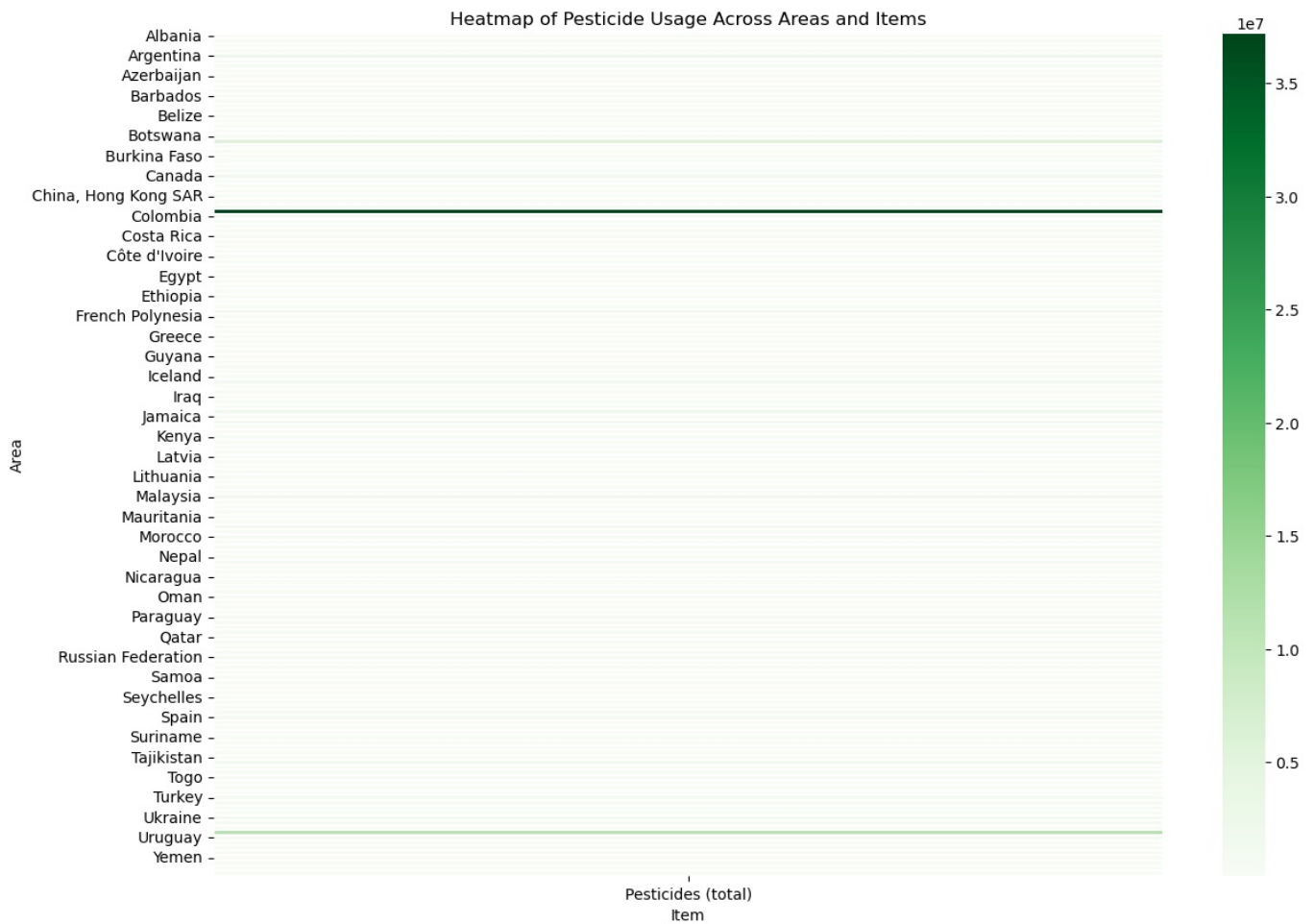
```
In [52]: # Creating a pivot table for heatmap
pivot_table = df.pivot_table(values='Value', index='Area', columns='Item', aggfunc='sum', fill_value=0)
print("Pivot table of pesticide usage across areas and items:\n", pivot_table)

# Plotting the heatmap
plt.figure(figsize=(14, 10))
sns.heatmap(pivot_table, cmap='Greens', linecolor='white', linewidths=0.1)
plt.title('Heatmap of Pesticide Usage Across Areas and Items')
plt.xlabel('Item')
plt.ylabel('Area')
plt.show()
```

Pivot table of pesticide usage across areas and items:

Item	Pesticides (total)
Area	
Albania	19509.24
Algeria	140318.85
Angola	1536.50
Antigua and Barbuda	337.73
Argentina	2250064.53
...	...
Viet Nam	555722.00
Yemen	16901.55
Yugoslav SFR	5432.00
Zambia	41186.00
Zimbabwe	83106.25

[168 rows x 1 columns]



Which years have seen the highest increase or decrease in pesticide usage?

```
In [55]: # Calculate the year-over-year change in pesticide usage
yearly_usage_change = yearly_usage.pct_change() * 100
print("Year-over-year change in pesticide usage (%):\n", yearly_usage_change)

# Plotting the year-over-year change
plt.figure(figsize=(12, 6))
yearly_usage_change.plot(kind='bar', color=green_color)
plt.title('Year-over-Year Change in Pesticide Usage (%)')
plt.xlabel('Year')
plt.ylabel('Change in Pesticide Usage (%)')
plt.grid(True)
plt.show()
```

Year-over-year change in pesticide usage (%):

Year

1990 NaN

1991 -1.118455

1992 2.878401

1993 2.566511

1994 6.661426

1995 5.851639

1996 4.280893

1997 3.885642

1998 1.756200

1999 3.468908

2000 -1.104813

2001 -1.312234

2002 0.792817

2003 2.944177

2004 5.694598

2005 2.107559

2006 1.098299

2007 7.843126

2008 1.110119

2009 -1.364222

2010 5.546561

2011 2.826740

2012 2.282373

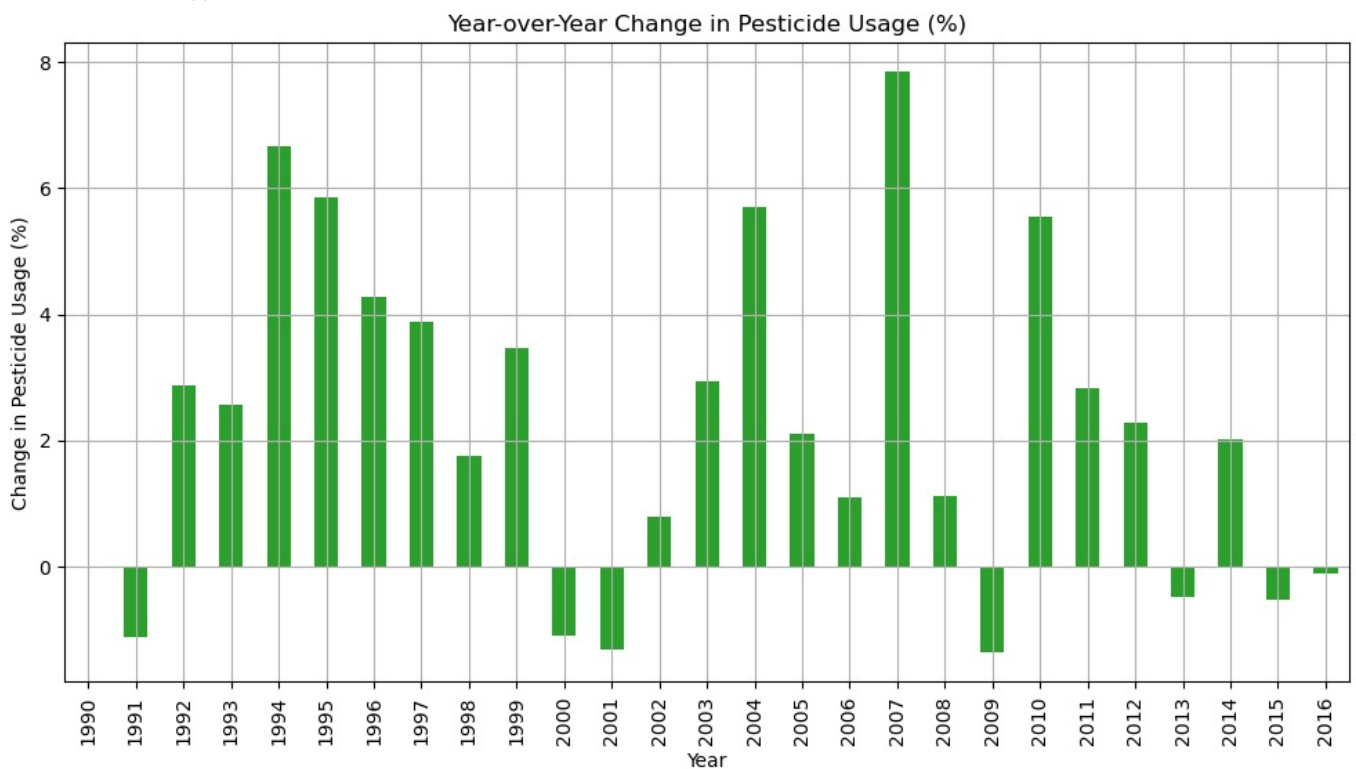
2013 -0.478404

2014 2.010182

2015 -0.530568

2016 -0.106466

Name: Value, dtype: float64



What is the average pesticide usage per area?

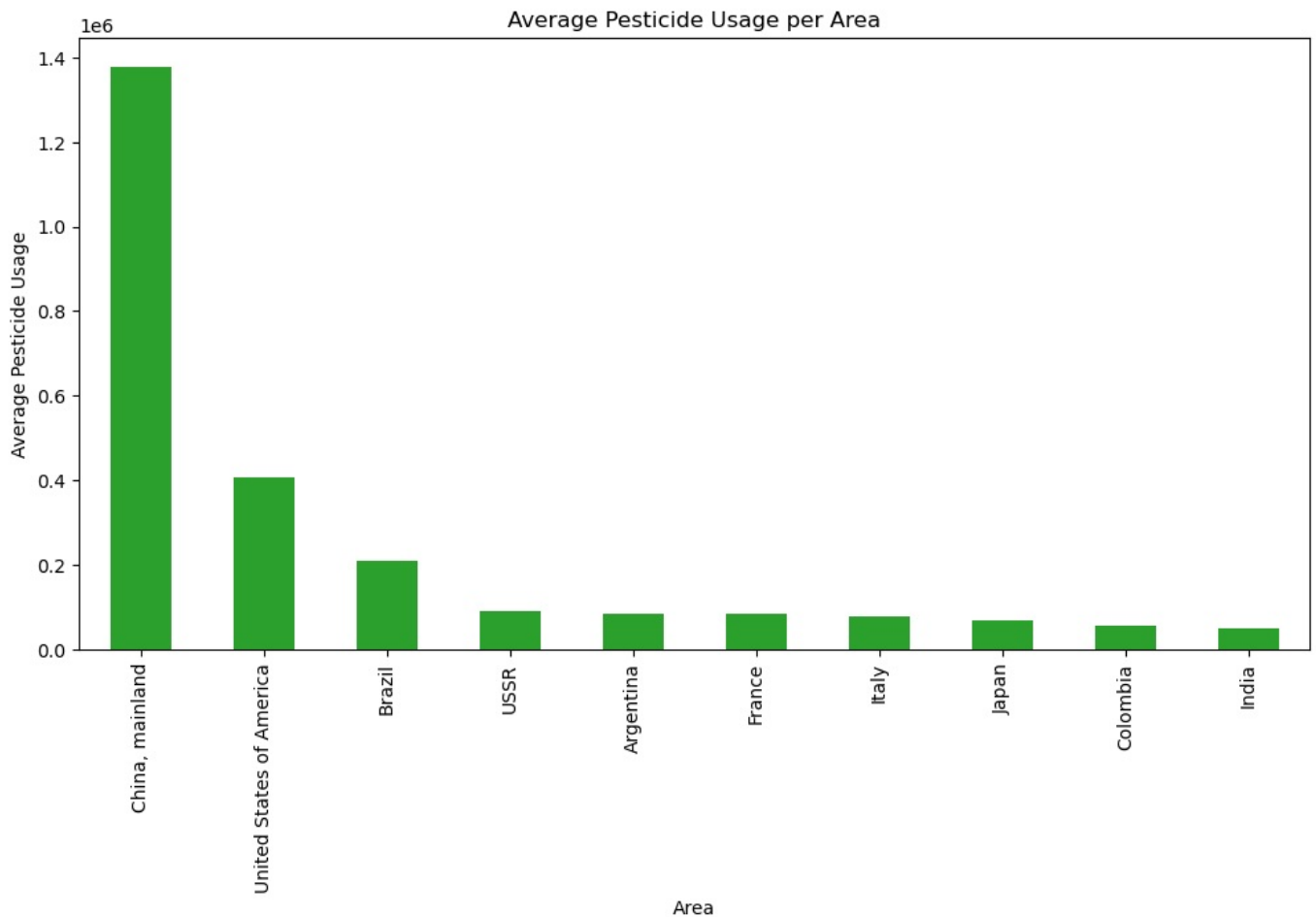
```
In [58]: # Calculate the average pesticide usage per area
average_area_usage = df.groupby('Area')['Value'].mean().sort_values(ascending=False)
print("Average pesticide usage per area:\n", average_area_usage.head(10))

# Plotting the average usage per area
plt.figure(figsize=(12, 6))
average_area_usage.head(10).plot(kind='bar', color=green_color)
plt.title('Average Pesticide Usage per Area')
plt.xlabel('Area')
plt.ylabel('Average Pesticide Usage')
plt.show()
```

Average pesticide usage per area:

Area	
China, mainland	1.377587e+06
United States of America	4.066033e+05
Brazil	2.100575e+05
USSR	8.920000e+04
Argentina	8.333572e+04
France	8.327432e+04
Italy	7.826321e+04
Japan	6.826507e+04
Colombia	5.486039e+04
India	4.898567e+04

Name: Value, dtype: float64



Is there any seasonality in pesticide usage over the years?

```
In [72]: # Assuming the data has monthly information, extract the month and calculate monthly averages
df['Month'] = pd.to_datetime(df['Year'], format='%Y').dt.month
monthly_usage = df.groupby('Month')['Value'].mean()
print("Average pesticide usage by month:\n", monthly_usage)

# Plotting the monthly usage
plt.figure(figsize=(12, 6))
monthly_usage.plot(kind='line', marker='o', color=green_color)
plt.title('Average Pesticide Usage by Month')
plt.xlabel('Month')
plt.ylabel('Average Pesticide Usage')
plt.grid(True)
plt.show()
```

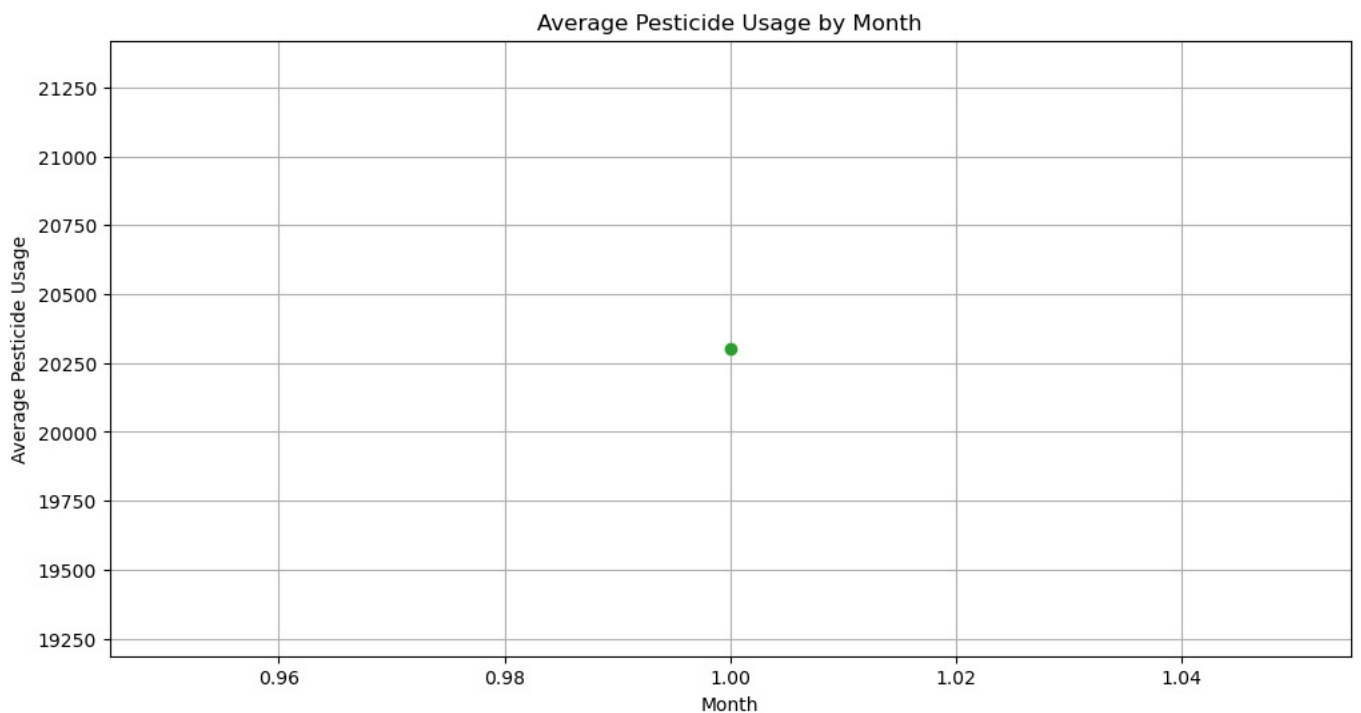
Average pesticide usage by month:

Month

1 20303.341465

Name: Value, dtype: float64





Which combination of area and element has the highest pesticide usage?

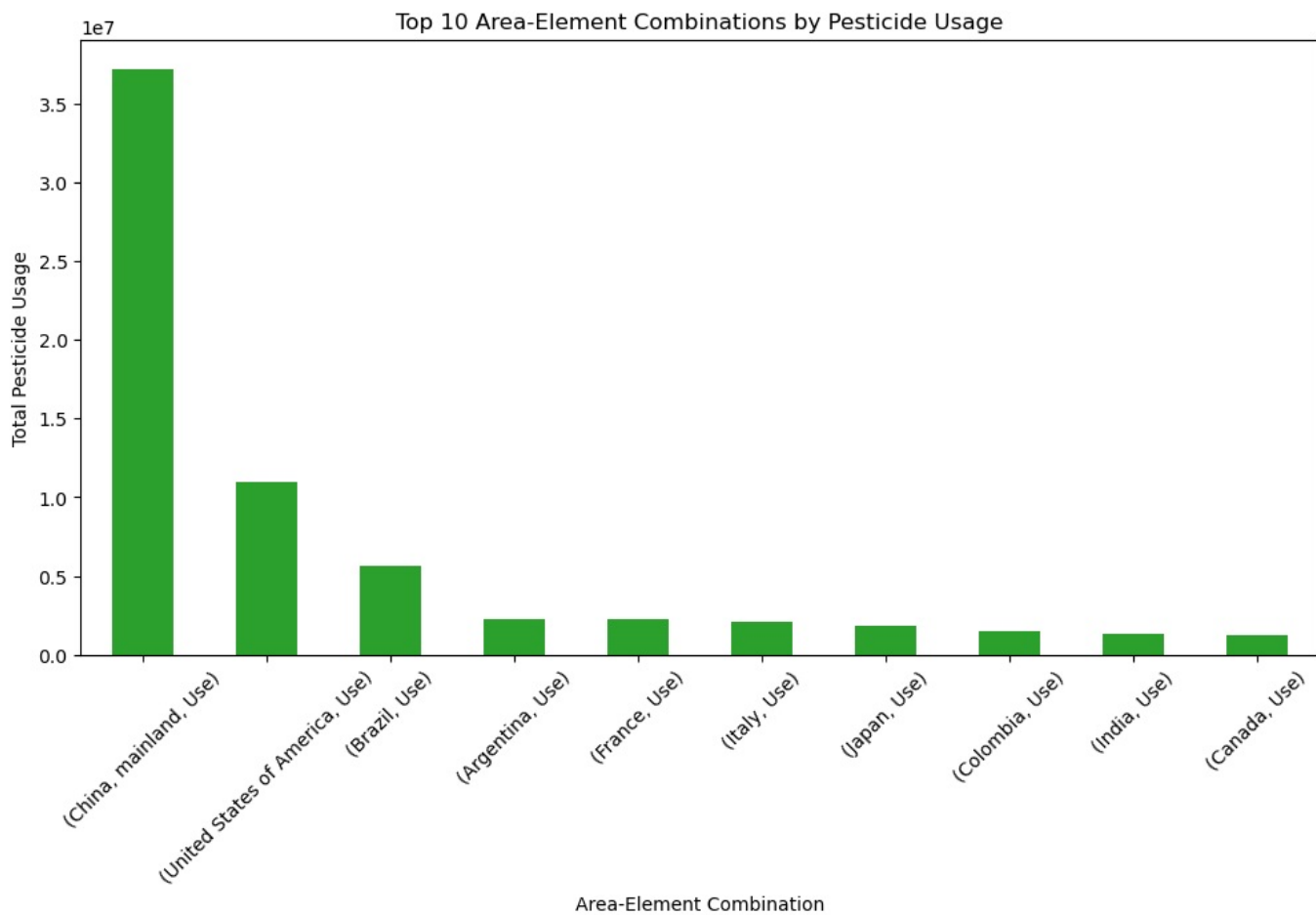
```
In [76]: # Calculate the combination of area and element with the highest pesticide usage
area_element_usage = df.groupby(['Area', 'Element'])['Value'].sum().sort_values(ascending=False)
print("Top area-element combinations by pesticide usage:\n", area_element_usage.head(10))

# Plotting the top 10 area-element combinations
plt.figure(figsize=(12, 6))
area_element_usage.head(10).plot(kind='bar', color=green_color)
plt.title('Top 10 Area-Element Combinations by Pesticide Usage')
plt.xlabel('Area-Element Combination')
plt.ylabel('Total Pesticide Usage')
plt.xticks(rotation=45)
plt.show()
```

Top area-element combinations by pesticide usage:

Area	Element	Value
China, mainland	Use	37194836.00
United States of America	Use	10978289.10
Brazil	Use	5671552.00
Argentina	Use	2250064.53
France	Use	2248406.72
Italy	Use	2113106.60
Japan	Use	1843156.79
Colombia	Use	1481230.51
India	Use	1322613.00
Canada	Use	1253897.98

Name: Value, dtype: float64



In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js