```
from google.colab import files
uploaded=files.upload()
print(uploaded)
```

Choose Files SupplyChai…odities.xlsx

- **SupplyChainEmissionFactorsforUSIndustriesCommodities.xlsx**(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 1416112 bytes, last modified: 6/16/2025 - 100% done

```
Saving SupplyChainEmissionFactorsforUSIndustriesCommodities.xlsx to SupplyChainEmissionFactorsforUSIndustriesCommodities.xlsx
{'SupplyChainEmissionFactorsforUSIndustriesCommodities.xlsx': b'PK\x03\x04\x14\x00\x00\x00\x08\x00\x00\x00!\x00\x85F\xa6~\x05\x02\x00\x0
```

Step 1: Import Required Libraries

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
import joblib
```

Step 2: Load Dataset

```
excel_file = 'SupplyChainEmissionFactorsforUSIndustriesCommodities.xlsx'
years = range(2010, 2017)
```

```
years[2]
```

```
2012
```

```
df_1 = pd.read_excel(excel_file, sheet_name=f'{years[0]}_Detail_Commodity')
df_1.head()
```

	Commodity Code	Commodity Name	Substance	Unit	Supply Chain Emission Factors without Margins	Margins of Supply Chain Emission Factors	Supply Chain Emission Factors with Margins	Unnamed: 7	DQ ReliabilityScore of Factors without Margins	DQ TemporalCorrelation of Factors without Margins	Geographic of Fa
0	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	carbon dioxide	kg/2018 USD, purchaser price	0.398	0.073	0.470	NaN	4	3	
1	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	methane	kg/2018 USD, purchaser price	0.001	0.001	0.002	NaN	4	3	
2	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	nitrous oxide	kg/2018 USD, purchaser price	0.002	0.000	0.002	NaN	4	3	
3	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	other GHGs	kg CO2e/2018 USD, purchaser price	0.002	0.000	0.002	NaN	3	3	
4	1111B0	Fresh wheat, corn, rice, and other grains	carbon dioxide	kg/2018 USD, purchaser price	0.659	0.081	0.740	NaN	4	3	

Next steps: Generate code with df_1 View recommended plots New interactive sheet

```
df_2 = pd.read_excel(excel_file, sheet_name=f'{years[0]}_Detail_Industry')
df_2.head()
```

	Industry Code	Industry Name	Substance	Unit	Supply Chain Emission Factors without Margins	Margins of Supply Chain Emission Factors	Supply Chain Emission Factors with Margins	Unnamed: 7	DQ ReliabilityScore of Factors without Margins	DQ TemporalCorrelation of Factors without Margins	Geographical of Facto
0	1111A0	Oilseed farming	carbon dioxide	kg/2018 USD, purchaser price	0.414	0.073	0.487	NaN	4	3	
1	1111A0	Oilseed farming	methane	kg/2018 USD, purchaser price	0.001	0.001	0.002	NaN	4	3	
2	1111A0	Oilseed farming	nitrous oxide	kg/2018 USD, purchaser price	0.002	0.000	0.002	NaN	4	3	
3	1111A0	Oilseed farming	other GHGs	kg CO2e/2018 USD, purchaser price	0.002	0.000	0.002	NaN	3	3	
4	1111B0	Grain farming	carbon dioxide	kg/2018 USD, purchaser price	0.680	0.082	0.762	NaN	4	3	

Next steps: Generate code with df_2 View recommended plots New interactive sheet

```
all_data = []

for year in years:
    try:
        df_com = pd.read_excel(excel_file, sheet_name=f'{year}_Detail_Commodity')
        df_ind = pd.read_excel(excel_file, sheet_name=f'{year}_Detail_Industry')

        df_com['Source'] = 'Commodity'
        df_ind['Source'] = 'Industry'
        df_com['Year'] = df_ind['Year'] = year

        df_com.columns = df_com.columns.str.strip()
        df_ind.columns = df_ind.columns.str.strip()

        df_com.rename(columns={
            'Commodity Code': 'Code',
            'Commodity Name': 'Name'
        }, inplace=True)

        df_ind.rename(columns={
            'Industry Code': 'Code',
            'Industry Name': 'Name'
        }, inplace=True)

        all_data.append(pd.concat([df_com, df_ind], ignore_index=True))

    except Exception as e:
        print(f"Error processing year {year}: {e}")


all_data[3]
```

	Code	Name	Substance	Unit	Supply Chain Emission Factors without Margins	Margins of Supply Chain Emission Factors	Supply Chain Emission Factors with Margins	Unnamed: 7	DQ ReliabilityScore of Factors without Margins	DQ TemporalCorrelation of Factors without Margins	Geograp of
0	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	carbon dioxide	kg/2018 USD, purchaser price	0.373	0.072	0.444	NaN	4	3	
1	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	methane	kg/2018 USD, purchaser price	0.001	0.001	0.002	NaN	4	3	
2	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	nitrous oxide	kg/2018 USD, purchaser price	0.002	0.000	0.002	NaN	4	3	
3	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	other GHGs	kg CO2e/2018 USD, purchaser price	0.002	0.000	0.002	NaN	3	3	
4	1111B0	Fresh wheat, corn, rice, and other grains	carbon dioxide	kg/2018 USD, purchaser price	0.722	0.079	0.801	NaN	4	3	
...	...	...	...	...	...	...	...	...	...	...	...
3151	813B00	Civic, social, professional, and similar organ...	other GHGs	kg CO2e/2018 USD, purchaser price	0.008	0.000	0.008	NaN	4	3	
3152	814000	Private households	carbon dioxide	kg/2018 USD, purchaser price	0.000	0.000	0.000	NaN	4	3	
3153	814000	Private households	methane	kg/2018 USD, purchaser price	0.000	0.000	0.000	NaN	4	3	
3154	814000	Private households	nitrous oxide	kg/2018 USD, purchaser price	0.000	0.000	0.000	NaN	4	3	
3155	814000	Private households	other GHGs	kg CO2e/2018 USD, purchaser price	0.000	0.000	0.000	NaN	4	3	

3156 rows × 15 columns

```
len(all_data)
```

```
7
```

```
df = pd.concat(all_data, ignore_index=True)
df.head(10)
```

	Code	Name	Substance	Unit	Supply Chain Emission Factors without Margins	Margins of Supply Chain Emission Factors	Supply Chain Emission Factors with Margins	Unnamed: 7	DQ ReliabilityScore of Factors without Margins	DQ TemporalCorrelation of Factors without Margins	Geographica of Fac
0	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	carbon dioxide	kg/2018 USD, purchaser price	0.398	0.073	0.470	NaN	4	3	
1	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	methane	kg/2018 USD, purchaser price	0.001	0.001	0.002	NaN	4	3	
2	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	nitrous oxide	kg/2018 USD, purchaser price	0.002	0.000	0.002	NaN	4	3	
3	1111A0	Fresh soybeans, canola, flaxseeds, and other o...	other GHGs	kg CO2e/2018 USD, purchaser price	0.002	0.000	0.002	NaN	3	3	
4	1111B0	Fresh wheat, corn, rice, and other grains	carbon dioxide	kg/2018 USD, purchaser price	0.659	0.081	0.740	NaN	4	3	
5	1111B0	Fresh wheat, corn, rice, and other grains	methane	kg/2018 USD, purchaser price	0.008	0.001	0.009	NaN	2	3	
6	1111B0	Fresh wheat, corn, rice, and other grains	nitrous oxide	kg/2018 USD, purchaser price	0.004	0.000	0.004	NaN	4	3	
7	1111B0	Fresh wheat, corn, rice, and other grains	other GHGs	kg CO2e/2018 USD, purchaser price	0.004	0.000	0.004	NaN	3	3	
8	111200	Fresh vegetables, melons, and potatoes	carbon dioxide	kg/2018 USD, purchaser price	0.183	0.132	0.315	NaN	3	3	