



Week 2

```
In [1]: # As there is no data available in Unnamed column so we will drop the column  
df.drop(columns=['Unnamed: 7'],inplace=True)
```

```
-----  
NameError                                Traceback (most recent call last)  
Cell In[1], line 2  
    1 # As there is no data available in Unnamed column so we will drop the c  
column  
----> 2 df.drop(columns=['Unnamed: 7'],inplace=True)  
  
NameError: name 'df' is not defined
```

```
In [92]: df.columns
```

```
Out[92]: Index(['Code', 'Name', 'Substance', 'Unit',  
               'Supply Chain Emission Factors without Margins',  
               'Margins of Supply Chain Emission Factors',  
               'Supply Chain Emission Factors with Margins',  
               'DQ ReliabilityScore of Factors without Margins',  
               'DQ TemporalCorrelation of Factors without Margins',  
               'DQ GeographicalCorrelation of Factors without Margins',  
               'DQ TechnologicalCorrelation of Factors without Margins',  
               'DQ DataCollection of Factors without Margins', 'Source', 'Year'],  
              dtype='object')
```

```
In [93]: print(df.info())    # Checking data types and non-null counts
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22092 entries, 0 to 22091
Data columns (total 14 columns):
#   Column                                                                 Non-Null Count  Dt
type  -----
-----
0    Code                                                                22092 non-null  ob
ject
1    Name                                                                22092 non-null  ob
ject
2    Substance                                                            22092 non-null  ob
ject
3    Unit                                                                22092 non-null  ob
ject
4    Supply Chain Emission Factors without Margins                    22092 non-null  fl
oat64
5    Margins of Supply Chain Emission Factors                        22092 non-null  fl
oat64
6    Supply Chain Emission Factors with Margins                      22092 non-null  fl
oat64
7    DQ ReliabilityScore of Factors without Margins                  22092 non-null  in
t64
8    DQ TemporalCorrelation of Factors without Margins              22092 non-null  in
t64
9    DQ GeographicalCorrelation of Factors without Margins          22092 non-null  in
t64
10   DQ TechnologicalCorrelation of Factors without Margins          22092 non-null  in
t64
11   DQ DataCollection of Factors without Margins                   22092 non-null  in
t64
12   Source                                                            22092 non-null  ob
ject
13   Year                                                            22092 non-null  in
t64
dtypes: float64(3), int64(6), object(5)
memory usage: 2.4+ MB
None

```

```
In [94]: df.describe().T # Checking summary statistics
```

Out[94]:

	count	mean	std	min	25%	50%
Supply Chain Emission Factors without Margins	22092.0	0.084807	0.267039	0.0	0.0	0.00
Margins of Supply Chain Emission Factors	22092.0	0.012857	0.078720	0.0	0.0	0.00
Supply Chain Emission Factors with Margins	22092.0	0.097681	0.288992	0.0	0.0	0.00
DQ ReliabilityScore of Factors without Margins	22092.0	3.308030	0.499643	2.0	3.0	3.00
DQ TemporalCorrelation of Factors without Margins	22092.0	2.571429	0.494883	2.0	2.0	3.00
DQ GeographicalCorrelation of Factors without Margins	22092.0	1.000000	0.000000	1.0	1.0	1.00
DQ TechnologicalCorrelation of Factors without Margins	22092.0	2.632129	1.135661	1.0	1.0	3.00
DQ DataCollection of Factors without Margins	22092.0	1.000000	0.000000	1.0	1.0	1.00
Year	22092.0	2013.000000	2.000045	2010.0	2011.0	2013.00

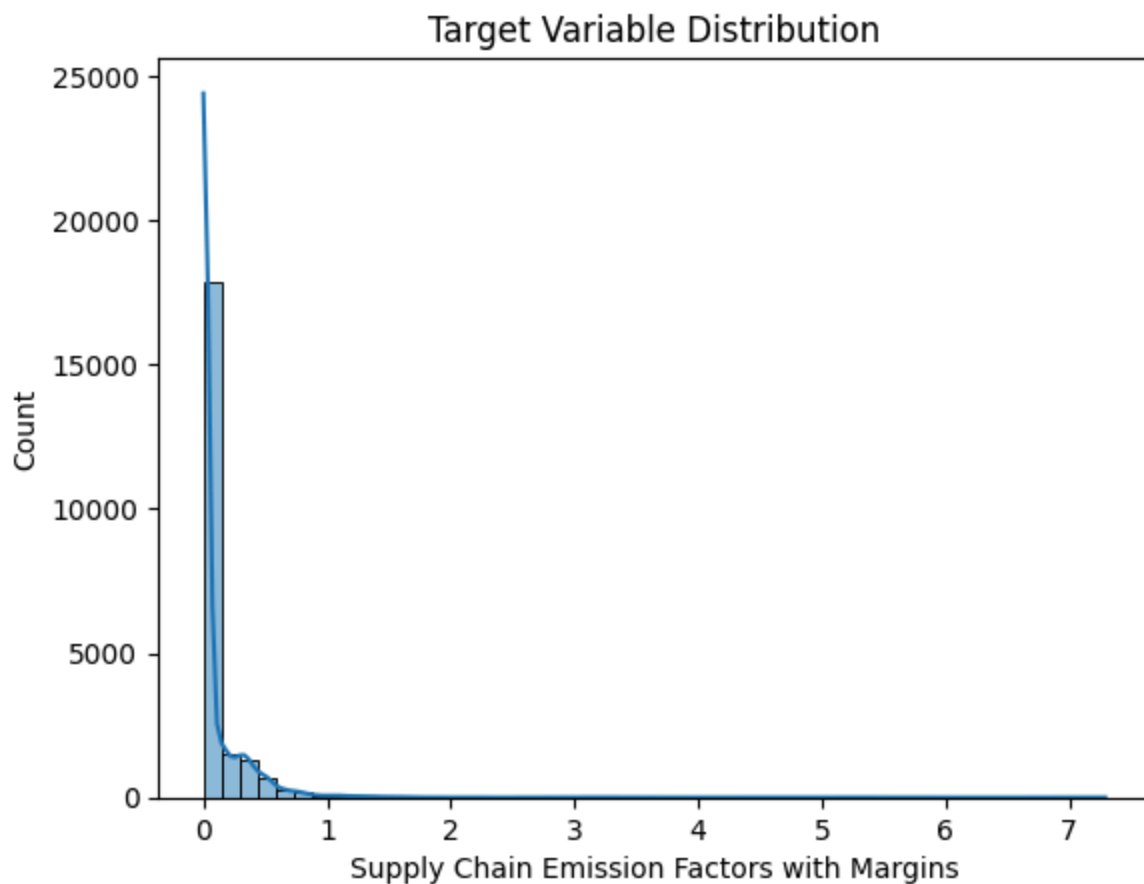
In [95]: `df.isnull().sum()` # Checking for null values in each column

Out[95]:

	0
Code	0
Name	0
Substance	0
Unit	0
Supply Chain Emission Factors without Margins	0
Margins of Supply Chain Emission Factors	0
Supply Chain Emission Factors with Margins	0
DQ ReliabilityScore of Factors without Margins	0
DQ TemporalCorrelation of Factors without Margins	0
DQ GeographicalCorrelation of Factors without Margins	0
DQ TechnologicalCorrelation of Factors without Margins	0
DQ DataCollection of Factors without Margins	0
Source	0
Year	0

dtype: int64

```
In [96]: # Visualize distribution
sns.histplot(df['Supply Chain Emission Factors with Margins'], bins=50, kde=True)
plt.title('Target Variable Distribution')
plt.show()
```



```
In [97]: # Check categorical variables
print(df['Substance'].value_counts())
```

```
Substance
carbon dioxide    5523
methane           5523
nitrous oxide     5523
other GHGs        5523
Name: count, dtype: int64
```

```
In [98]: print(df['Unit'].value_counts()) # Checking unique values in 'Unit' with count
```

```
Unit
kg/2018 USD, purchaser price    16569
kg C02e/2018 USD, purchaser price  5523
Name: count, dtype: int64
```

```
In [99]: print(df['Unit'].unique()) # Checking unique values in 'Unit'
```

```
['kg/2018 USD, purchaser price' 'kg C02e/2018 USD, purchaser price']
```

```
In [100]: print(df['Source'].value_counts()) # Checking unique values in 'Source' with count
```

```
Source
Industry    11060
Commodity    11032
Name: count, dtype: int64
```

```
In [101... df['Substance'].unique() # Checking unique values in 'Substance'
Out[101... array(['carbon dioxide', 'methane', 'nitrous oxide', 'other GHGs'],
              dtype=object)
In [102... substance_map={'carbon dioxide':0, 'methane':1, 'nitrous oxide':2, 'other GHGs':3}
In [103... df['Substance']=df['Substance'].map(substance_map)
In [104... df['Substance'].unique() # Checking unique values in 'Substance'
Out[104... array([0, 1, 2, 3])
In [105... print(df['Unit'].unique()) # Checking unique values in 'Unit'
          ['kg/2018 USD, purchaser price' 'kg CO2e/2018 USD, purchaser price']
In [106... unit_map={'kg/2018 USD, purchaser price':0, 'kg CO2e/2018 USD, purchaser price':1}
In [107... df['Unit']=df['Unit'].map(unit_map)
In [108... print(df['Unit'].unique()) # Checking unique values in 'Unit'
          [0 1]
In [109... print(df['Source'].unique()) # Checking unique values in 'Source'
          ['Commodity' 'Industry']
In [110... source_map={'Commodity':0, 'Industry':1} # Mapping sources to integers
In [111... df['Source']=df['Source'].map(source_map) # applying the mapping to 'Source'
In [112... print(df['Source'].unique()) # Checking unique values in 'Source'
          [0 1]
In [113... df.info() # Checking data types and non-null counts after mapping
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22092 entries, 0 to 22091
Data columns (total 14 columns):
#   Column                                                                 Non-Null Count  Dtype
---  -
0   Code                                                                    22092 non-null  object
1   Name                                                                    22092 non-null  object
2   Substance                                                                22092 non-null  int64
3   Unit                                                                    22092 non-null  int64
4   Supply Chain Emission Factors without Margins                        22092 non-null  float64
5   Margins of Supply Chain Emission Factors                            22092 non-null  float64
6   Supply Chain Emission Factors with Margins                          22092 non-null  float64
7   DQ ReliabilityScore of Factors without Margins                      22092 non-null  int64
8   DQ TemporalCorrelation of Factors without Margins                  22092 non-null  int64
9   DQ GeographicalCorrelation of Factors without Margins              22092 non-null  int64
10  DQ TechnologicalCorrelation of Factors without Margins              22092 non-null  int64
11  DQ DataCollection of Factors without Margins                        22092 non-null  int64
12  Source                                                                    22092 non-null  int64
13  Year                                                                    22092 non-null  int64
dtypes: float64(3), int64(9), object(2)
memory usage: 2.4+ MB

```

```
In [114... df.Code.unique() # Checking unique values in 'Code' df['Code']
```

```
Out[114]: array(['1111A0', '1111B0', '111200', '111300', '111400', '111900',
'112120', '1121A0', '112300', '112A00', '113000', '114000',
'115000', '211000', '212100', '212230', '2122A0', '212310',
'2123A0', '213111', '21311A', '221100', '221200', '221300',
'230301', '230302', '233210', '233230', '233240', '233262',
'2332A0', '2332C0', '2332D0', '233411', '233412', '2334A0',
'311111', '311119', '311210', '311221', '311224', '311225',
'311230', '311300', '311410', '311420', '311513', '311514',
'31151A', '311520', '311615', '31161A', '311700', '311810',
'3118A0', '311910', '311920', '311930', '311940', '311990',
'312110', '312120', '312130', '312140', '312200', '313100',
'313200', '313300', '314110', '314120', '314900', '315000',
'316000', '321100', '321200', '321910', '3219A0', '322110',
'322120', '322130', '322210', '322220', '322230', '322291',
'322299', '323110', '323120', '324110', '324121', '324122',
'324190', '325110', '325120', '325130', '325180', '325190',
'325211', '3252A0', '325310', '325320', '325411', '325412',
'325413', '325414', '325510', '325520', '325610', '325620',
'325910', '3259A0', '326110', '326120', '326130', '326140',
'326150', '326160', '326190', '326210', '326220', '326290',
'327100', '327200', '327310', '327320', '327330', '327390',
'327400', '327910', '327991', '327992', '327993', '327999',
'331110', '331200', '331313', '33131B', '331410', '331420',
'331490', '331510', '331520', '332114', '332119', '33211A',
'332200', '332310', '332320', '332410', '332420', '332430',
'332500', '332600', '332710', '332720', '332800', '332913',
'33291A', '332991', '332996', '332999', '33299A', '333111',
'333112', '333120', '333130', '333242', '33329A', '333314',
'333316', '333318', '333413', '333414', '333415', '333511',
'333514', '333517', '33351B', '333611', '333612', '333613',
'333618', '333912', '33391A', '333920', '333991', '333993',
'333994', '33399A', '33399B', '334111', '334112', '334118',
'334210', '334220', '334290', '334300', '334413', '334418',
'33441A', '334510', '334511', '334512', '334513', '334514',
'334515', '334516', '334517', '33451A', '334610', '335110',
'335120', '335210', '335221', '335222', '335224', '335228',
'335311', '335312', '335313', '335314', '335911', '335912',
'335920', '335930', '335991', '335999', '336111', '336112',
'336120', '336211', '336212', '336213', '336214', '336310',
'336320', '336350', '336360', '336370', '336390', '3363A0',
'336411', '336412', '336413', '336414', '33641A', '336500',
'336611', '336612', '336991', '336992', '336999', '337110',
'337121', '337122', '337127', '33712N', '337215', '33721A',
'337900', '339112', '339113', '339114', '339115', '339116',
'339910', '339920', '339930', '339940', '339950', '339990',
'4200ID', '423100', '423400', '423600', '423800', '423A00',
'424200', '424400', '424700', '424A00', '425000', '441000',
'444000', '445000', '446000', '447000', '448000', '452000',
'454000', '481000', '482000', '483000', '484000', '485000',
'486000', '48A000', '491000', '492000', '493000', '4B0000',
'511110', '511120', '511130', '5111A0', '511200', '512100',
'512200', '515100', '515200', '517110', '517210', '517A00',
'518200', '519130', '5191A0', '522A00', '523900', '523A00',
'524113', '5241XX', '524200', '525000', '52A000', '531HS0',
```



```
'531HST', '5310RE', '532100', '532400', '532A00', '533000',  
'541100', '541200', '541300', '541400', '541511', '541512',  
'54151A', '541610', '5416A0', '541700', '541800', '541920',  
'541940', '5419A0', '550000', '561100', '561200', '561300',  
'561400', '561500', '561600', '561700', '561900', '562000',  
'611100', '611A00', '611B00', '621100', '621200', '621300',  
'621400', '621500', '621600', '621900', '622000', '623A00',  
'623B00', '624100', '624400', '624A00', '711100', '711200',  
'711500', '711A00', '712000', '713100', '713200', '713900',  
'721000', '722110', '722211', '722A00', '811100', '811200',  
'811300', '811400', '812100', '812200', '812300', '812900',  
'813100', '813A00', '813B00', '814000', '331314'], dtype=object)
```

```
In [115... df.Name.unique() # Checking unique values in 'Name'
```

```
Out[115... array(['Fresh soybeans, canola, flaxseeds, and other oilseeds',
                  'Fresh wheat, corn, rice, and other grains',
                  'Fresh vegetables, melons, and potatoes',
                  'Fresh fruits and tree nuts',
                  'Greenhouse crops, mushrooms, nurseries, and flowers',
                  'Tobacco, cotton, sugarcane, peanuts, sugar beets, herbs and spices, a
nd other crops',
                  'Dairies', 'Cattle ranches and feedlots', 'Poultry farms',
                  'Animal farms and aquaculture ponds (except cattle and poultry)',
                  'Timber and raw forest products', 'Wild-caught fish and game',
                  'Agriculture and forestry support', 'Unrefined oil and gas',
                  'Coal', 'Copper, nickel, lead, and zinc',
                  'Iron, gold, silver, and other metal ores', 'Dimensional stone',
                  'Sand, gravel, clay, phosphate, other nonmetallic minerals',
                  'Well drilling', 'Other support activities for mining',
                  'Electricity', 'Natural gas',
                  'Drinking water and wastewater treatment',
                  'Nonresidential maintenance and repair',
                  'Residential maintenance and repair', 'Health care structures',
                  'Manufacturing structures', 'Power and communication structures',
                  'Educational and vocational structures',
                  'Commercial structures, including farm structures',
                  'Transportation structures and highways and streets',
                  'Other nonresidential structures',
                  'Single-family residential structures',
                  'Multifamily residential structures',
                  'Other residential structures', 'Dog and cat food',
                  'Other animal food', 'Flours and malts', 'Corn products',
                  'Soybean and other oilseed processing',
                  'Refined vegetable, olive, and seed oils', 'Breakfast cereals',
                  'Sugar, candy, and chocolate', 'Frozen food',
                  'Fruit and vegetable preservation', 'Cheese',
                  'Dry, condensed, and evaporated dairy', 'Fluid milk and butter',
                  'Ice cream and frozen desserts', 'Packaged poultry',
                  'Packaged meat (except poultry)', 'Seafood',
                  'Bread and other baked goods',
                  'Cookies, crackers, pastas, and tortillas', 'Snack foods',
                  'Coffee and tea', 'Flavored drink concentrates',
                  'Seasonings and dressings', 'All other foods',
                  'Soft drinks, bottled water, and ice', 'Breweries and beer',
                  'Wineries and wine', 'Distilleries and spirits',
                  'Tobacco products', 'Fiber, yarn, and thread', 'Fabric',
                  'Finished and coated fabric', 'Carpets and rugs',
                  'Curtains and linens', 'Other textiles', 'Clothing', 'Leather',
                  'Lumber and treated lumber', 'Plywood and veneer',
                  'Wooden windows, door, and flooring',
                  'Veneer, plywood, and engineered wood', 'Wood pulp', 'Paper',
                  'Cardboard', 'Cardboard containers', 'Paper bags and coated paper',
                  'Stationery', 'Sanitary paper (tissues, napkins, diapers, etc.)',
                  'All other converted paper products',
                  'Books, newspapers, magazines, and other print media',
                  'Printing support',
                  'Gasoline, fuels, and by-products of petroleum refining',
                  'Asphalt pavement', 'Asphalt shingles',
```

- 'Other petroleum and coal products', 'Petrochemicals',
- 'Compressed Gases', 'Synthetic dyes and pigments',
- 'Other basic inorganic chemicals', 'Other basic organic chemicals',
- 'Plastics', 'Synthetic rubber and artificial and synthetic fibers',
- 'Fertilizers', 'Pesticides', 'Medicinal and botanical ingredients',
- 'Pharmaceutical products (pills, powders, solutions, etc.)',
- 'Blood sugar, pregnancy, and other diagnostic test kits',
- 'Vaccines and other biological medical products',
- 'Paints and coatings', 'Adhesives', 'Soap and cleaning compounds',
- 'Toiletries', 'Ink and ink cartridges',
- 'Chemicals (except basic chemicals, agrichemicals, polymers, paints, pharmaceuticals, soaps, cleaning compounds)',
- 'Plastic bags, films, and sheets',
- 'Plastic pipe, fittings, and sausage casings',
- 'Laminated plastic plates and shapes', 'Polystyrene foam products',
- 'Urethane and other foam products', 'Plastic bottles',
- 'Other plastic products', 'Rubber tires',
- 'Rubber and plastic belts and hoses', 'Other rubber products',
- 'Clay and ceramic products', 'Glass and glass products', 'Cement',
- 'Ready-mix concrete', 'Concrete pipe, bricks, and blocks',
- 'Other concrete products', 'Lime and gypsum products',
- 'Abrasive products', 'Cut stone and stone products',
- 'Ground or treated minerals and earth', 'Mineral wool',
- 'Other nonmetallic mineral products',
- 'Primary iron, steel, and ferroalloy products',
- 'Secondary steel products',
- 'Alumina refining and primary aluminum production',
- 'Aluminum products',
- 'Nonferrous Metal (except Aluminum) Smelting and Refining',
- 'Secondary copper products',
- 'Other secondary nonferrous metal products', 'Cast iron and steel',
- 'Nonferrous metal casts', 'Custom metal rolls',
- 'Metal crown, closure, and other metal stamping (except automotive)',
- 'All other forging, stamping, and sintering',
- 'Cutlery and handtools', 'Metal structural products',
- 'Metal windows, doors, and architectural products',
- 'Power boilers and heat exchangers', 'Heavy gauge metal tanks',
- 'Light gauge metal cans, boxes, and containers',
- 'Metal hinges, keys, lock, and other hardware',
- 'Springs and wires', 'Machine shops', 'Screws, nuts, and bolts',
- 'Metal coatings, engravings, and heat treatments',
- 'Metal plumbing drains, faucets, valves, and other fittings',
- 'Valve and fittings (except for plumbing)',
- 'Ball and roller bearings', 'Fabricated pipe and pipe fittings',
- 'Other fabricated metal manufacturing',
- 'Ammunition, arms, ordnance, and related accessories',
- 'Farm machinery and equipment', 'Lawn and garden equipment',
- 'Construction machinery', 'Mining and oil/gas field machinery',
- 'Semiconductor machinery',
- 'Machinery for the paper, textile, food or other industries (except semiconductor machinery)',
- 'Optical instruments and lenses',
- 'Photography and photocopying equipment',
- 'Other commercial and service industry machinery',

'Industrial and commercial fan and blower and air purification equipment',
 'Heating equipment other than warm air furnaces',
 'Air conditioning, refrigeration, and warm air heating equipment',
 'Industrial molds', 'Special tools, dies, jigs, and fixtures',
 'Machine tool manufacturing',
 'Cutting and machine tool accessory, rolling mill, and other metalworking machines',
 'Turbines and turbine generator sets',
 'Speed changers, industrial high-speed drives, and gears',
 'Mechanical power transmission equipment',
 'Other engine equipment', 'Air and gas compressors',
 'Pumps and pumping equipment', 'Material handling equipment',
 'Power tools', 'Packaging machinery',
 'Industrial process furnaces and ovens',
 'Welding and Soldering Equipment, Scales and Balances, and other general purpose machinery',
 'Hydraulic pumps, motors, cylinders and actuators', 'Computers',
 'Computer storage device readers',
 'Computer terminals and other computer peripheral equipment',
 'Telephones', 'Wireless communications',
 'Communications equipment', 'Audio and video equipment',
 'Semiconductors', 'Printed circuit and electronic assembly',
 'Electronic capacitors, resistors, coils, transformers, connectors and other components (except semiconductors and printed circuit assemblies)',
 'Electromedical apparatuses', 'Navigation instruments',
 'Automatic controls for HVAC and refrigeration equipment',
 'Industrial process variable instruments',
 'Fluid meters and counting devices', 'Signal testing instruments',
 'Analytical laboratory instruments', 'Irradiation apparatuses',
 'Watches, clocks, and other measuring and controlling devices',
 'External hard drives, CDs, other storage media', 'Light bulbs',
 'Light fixtures', 'Small electrical appliances',
 'Home cooking appliances', 'Home refrigerators and freezers',
 'Home laundry machines',
 'Major home appliances (except ovens, stoves, refrigerators and laundry machines)',
 'Specialty transformers', 'Motors and generators',
 'Switchgear and switchboards', 'Relay and industrial controls',
 'Storage batteries', 'Primary batteries',
 'Communication and energy wire and cable', 'Wiring devices',
 'Carbon and graphite products',
 'Other miscellaneous electrical equipment and components',
 'Automobiles', 'Pickup trucks, vans, and SUVs',
 'Heavy duty trucks', 'Vehicle bodies', 'Truck trailers',
 'Motor homes', 'Travel trailer and campers',
 'Vehicle engines and engine parts',
 'Vehicle electrical and electronic equipment',
 'Transmission and power train parts',
 'Vehicle seating and interior trim (upholstery)',
 'Vehicle metal stamping', 'Other vehicle parts',
 'Motor vehicle steering, suspension components (except spring), and brake systems',
 'Aircraft', 'Aircraft engines and parts', 'Other aircraft parts',

- 'Guided missiles and space vehicles',
- 'Propulsion units and parts for space vehicles and guided missiles',
- 'Railroad rolling stock', 'Ships and ship repair', 'Boats',
- 'Motorcycle, bicycle, and parts',
- 'Military armored vehicles and tanks',
- 'Other transportation equipment',
- 'Wood kitchen cabinets and countertops',
- 'Home furniture - upholstered',
- 'Home furniture - wood, nonupholstered', 'Institutional furniture',
- 'Other household nonupholstered furniture', 'Shelving and lockers',
- 'Office furniture and custom architectural woodwork and millwork',
- 'Mattresses, blinds and shades',
- 'Surgical and medical instruments',
- 'Surgical appliance and supplies', 'Dental equipment and supplies',
- 'Ophthalmic goods', 'Dental laboratories',
- 'Jewelry and silverware', 'Sporting and athletic goods',
- 'Dolls, toys, and games', 'Office supplies (not paper)', 'Signs',
- 'Gaskets, seals, musical instruments, fasteners, brooms, brushes, mop and other misc. goods',
- 'Customs duties',
- 'Motor vehicle and motor vehicle parts and supplies',
- 'Professional and commercial equipment and supplies',
- 'Household appliances and electrical and electronic goods',
- 'Machinery, equipment, and supplies',
- 'Other durable goods merchant wholesalers',
- 'Drugs and druggists' sundries',
- 'Grocery and related product wholesalers',
- 'Petroleum and petroleum products',
- 'Other nondurable goods merchant wholesalers',
- 'Wholesale electronic markets and agents and brokers',
- 'Vehicles and parts sales',
- 'Building material and garden equipment and supplies dealers',
- 'Food and beverage stores', 'Health and personal care stores',
- 'Gasoline stations', 'Clothing and clothing accessories stores',
- 'General merchandise stores', 'Nonstore retailers',
- 'Air transport', 'Rail transport',
- 'Water transport (boats, ships, ferries)', 'Truck transport',
- 'Passenger ground transport', 'Pipeline transport',
- 'Scenic and sightseeing transportation and support activities for transportation',
- 'Postal service', 'Couriers and messengers', 'Warehousing',
- 'All other retail', 'Newspapers', 'Magazines and journals',
- 'Books', 'Directory, mailing list, and other publishers',
- 'Software', 'Movies and film', 'Sound recording',
- 'Radio and television', 'Cable and subscription programming',
- 'Telecommunications', 'Wireless telecommunications',
- 'Satellite, telecommunications resellers, and all other telecommunications',
- 'Data processing and hosting',
- 'Internet publishing and broadcasting',
- 'News syndicates, libraries, archives, Internet publishing and all other information services',
- 'Nondepository credit intermediation and related activities',
- 'Investment advice, portfolio management, and other financial advising

services',
 'Securities and commodities brokerage and exchanges',
 'Direct life insurance carriers',
 'Insurance carriers, except direct life',
 'Insurance agencies and brokerages',
 'Funds, trusts, and financial vehicles',
 'Monetary authorities and depository credit intermediation',
 'Owner-occupied housing', 'Tenant-occupied housing',
 'Other real estate', 'Vehicle rental and leasing',
 'Commercial equipment rental',
 'Consumer goods and general rental centers',
 'Lessors of nonfinancial intangible assets', 'Legal services',
 'Accounting, tax preparation, bookkeeping, and payroll',
 'Architectural, engineering, and related services',
 'Specialized design', 'Custom computer programming',
 'Computer systems design',
 'Other computer related services, including facilities management',
 'Management consulting',
 'Environmental and other technical consulting services',
 'Scientific research and development',
 'Advertising and public relations', 'Photographers',
 'Veterinarians',
 'Marketing research and all other miscellaneous professional, scientific, and technical services',
 'Company and enterprise management', 'Office administration',
 'Facilities support', 'Employment services', 'Business support',
 'Travel arrangement and reservation', 'Investigation and security',
 'Buildings and dwellings services', 'Other support services',
 'Waste management and remediation',
 'Elementary and secondary schools',
 'Colleges, universities, junior colleges, and professional schools',
 'Other educational services', 'Physicians', 'Dentists',
 'Healthcare practitioners (except physicians and dentists)',
 'Outpatient healthcare', 'Medical laboratories', 'Home healthcare',
 'Ambulances', 'Hospitals', 'Nursing and community care facilities',
 'Residential mental retardation, mental health, substance abuse and other facilities',
 'Individual and family services', 'Child day care',
 'Community food, housing, and other relief services, including rehabilitation services',
 'Performances', 'Sports',
 'Independent artists, writers, and performers',
 'Promoters and agents',
 'Museums, historical sites, zoos, and parks',
 'Amusement parks and arcades',
 'Gambling establishments (except casino hotels)',
 'Golf courses, marinas, ski resorts, fitness and other rec centers and industries',
 'Hotels and campgrounds', 'Full-service restaurants',
 'Limited-service restaurants',
 'All other food and drinking places', 'Vehicle repair',
 'Electronic equipment repair and maintenance',
 'Commercial machinery repair', 'Household goods repair',
 'Salons and barber shops', 'Funerary services',

'Dry-cleaning and laundry',
'Pet care, photofinishing, parking and other sundry services',
'Religious organizations',
'Grantmaking, giving, and social advocacy organizations',
'Civic, social, professional, and similar organizations',
'Household employees', 'Oilseed farming', 'Grain farming',
'Vegetable and melon farming', 'Fruit and tree nut farming',
'Greenhouse, nursery, and floriculture production',
'Other crop farming', 'Dairy cattle and milk production',
'Beef cattle ranching and farming, including feedlots and dual-purpose
ranching and farming',
'Poultry and egg production',
'Animal production, except cattle and poultry and eggs',
'Forestry and logging', 'Fishing, hunting and trapping',
'Support activities for agriculture and forestry',
'Oil and gas extraction', 'Coal mining',
'Copper, nickel, lead, and zinc mining',
'Iron, gold, silver, and other metal ore mining',
'Stone mining and quarrying',
'Other nonmetallic mineral mining and quarrying',
'Drilling oil and gas wells',
'Electric power generation, transmission, and distribution',
'Natural gas distribution', 'Water, sewage and other systems',
'Office and commercial structures',
'Dog and cat food manufacturing',
'Other animal food manufacturing',
'Flour milling and malt manufacturing', 'Wet corn milling',
'Fats and oils refining and blending',
'Breakfast cereal manufacturing',
'Sugar and confectionery product manufacturing',
'Frozen food manufacturing',
'Fruit and vegetable canning, pickling, and drying',
'Cheese manufacturing',
'Dry, condensed, and evaporated dairy product manufacturing',
'Fluid milk and butter manufacturing',
'Ice cream and frozen dessert manufacturing', 'Poultry processing',
'Animal (except poultry) slaughtering, rendering, and processing',
'Seafood product preparation and packaging',
'Bread and bakery product manufacturing',
'Cookie, cracker, pasta, and tortilla manufacturing',
'Snack food manufacturing', 'Coffee and tea manufacturing',
'Flavoring syrup and concentrate manufacturing',
'Seasoning and dressing manufacturing',
'All other food manufacturing', 'Soft drink and ice manufacturing',
'Breweries', 'Wineries', 'Distilleries',
'Tobacco product manufacturing', 'Fiber, yarn, and thread mills',
'Fabric mills',
'Textile and fabric finishing and fabric coating mills',
'Carpet and rug mills', 'Curtain and linen mills',
'Other textile product mills', 'Apparel manufacturing',
'Leather and allied product manufacturing',
'Sawmills and wood preservation',
'Veneer, plywood, and engineered wood product manufacturing',
'Millwork', 'All other wood product manufacturing', 'Pulp mills',

'Paper mills', 'Paperboard mills',
 'Paperboard container manufacturing',
 'Paper Bag and Coated and Treated Paper Manufacturing',
 'Stationery product manufacturing',
 'Sanitary paper product manufacturing',
 'All other converted paper product manufacturing', 'Printing',
 'Support activities for printing', 'Petroleum refineries',
 'Asphalt paving mixture and block manufacturing',
 'Asphalt shingle and coating materials manufacturing',
 'Other petroleum and coal products manufacturing',
 'Petrochemical manufacturing', 'Industrial gas manufacturing',
 'Synthetic dye and pigment manufacturing',
 'Other Basic Inorganic Chemical Manufacturing',
 'Other basic organic chemical manufacturing',
 'Plastics material and resin manufacturing',
 'Synthetic rubber and artificial and synthetic fibers and filaments ma
 nufacturing',
 'Fertilizer manufacturing',
 'Pesticide and other agricultural chemical manufacturing',
 'Medicinal and botanical manufacturing',
 'Pharmaceutical preparation manufacturing',
 'In-vitro diagnostic substance manufacturing',
 'Biological product (except diagnostic) manufacturing',
 'Paint and coating manufacturing', 'Adhesive manufacturing',
 'Soap and cleaning compound manufacturing',
 'Toilet preparation manufacturing', 'Printing ink manufacturing',
 'All other chemical product and preparation manufacturing',
 'Plastics packaging materials and unlaminated film and sheet manufactu
 ring',
 'Plastics pipe, pipe fitting, and unlaminated profile shape manufactur
 ing',
 'Laminated plastics plate, sheet (except packaging), and shape manufac
 turing',
 'Polystyrene foam product manufacturing',
 'Urethane and other foam product (except polystyrene) manufacturing',
 'Plastics bottle manufacturing',
 'Other plastics product manufacturing', 'Tire manufacturing',
 'Rubber and plastics hoses and belting manufacturing',
 'Other rubber product manufacturing',
 'Clay product and refractory manufacturing',
 'Glass and glass product manufacturing', 'Cement manufacturing',
 'Ready-mix concrete manufacturing',
 'Concrete pipe, brick, and block manufacturing',
 'Other concrete product manufacturing',
 'Lime and gypsum product manufacturing',
 'Abrasive product manufacturing',
 'Cut stone and stone product manufacturing',
 'Ground or treated mineral and earth manufacturing',
 'Mineral wool manufacturing',
 'Miscellaneous nonmetallic mineral products',
 'Iron and steel mills and ferroalloy manufacturing',
 'Steel product manufacturing from purchased steel',
 'Secondary smelting and alloying of aluminum',
 'Aluminum product manufacturing from purchased aluminum',

- 'Copper rolling, drawing, extruding and alloying',
- 'Nonferrous metal (except copper and aluminum) rolling, drawing, extruding and alloying',
- 'Ferrous metal foundries', 'Nonferrous metal foundries',
- 'Custom roll forming', 'Cutlery and handtool manufacturing',
- 'Plate work and fabricated structural product manufacturing',
- 'Ornamental and architectural metal products manufacturing',
- 'Power boiler and heat exchanger manufacturing',
- 'Metal tank (heavy gauge) manufacturing',
- 'Metal can, box, and other metal container (light gauge) manufacturing',
- 'Hardware manufacturing', 'Spring and wire product manufacturing',
- 'Turned product and screw, nut, and bolt manufacturing',
- 'Coating, engraving, heat treating and allied activities',
- 'Plumbing fixture fitting and trim manufacturing',
- 'Valve and fittings other than plumbing',
- 'Ball and roller bearing manufacturing',
- 'Fabricated pipe and pipe fitting manufacturing',
- 'Ammunition, arms, ordnance, and accessories manufacturing',
- 'Farm machinery and equipment manufacturing',
- 'Lawn and garden equipment manufacturing',
- 'Construction machinery manufacturing',
- 'Mining and oil and gas field machinery manufacturing',
- 'Semiconductor machinery manufacturing',
- 'Other industrial machinery manufacturing',
- 'Optical instrument and lens manufacturing',
- 'Photographic and photocopying equipment manufacturing',
- 'Other commercial and service industry machinery manufacturing',
- 'Industrial and commercial fan and blower and air purification equipment manufacturing',
- 'Heating equipment (except warm air furnaces) manufacturing',
- 'Air conditioning, refrigeration, and warm air heating equipment manufacturing',
- 'Industrial mold manufacturing',
- 'Special tool, die, jig, and fixture manufacturing',
- 'Cutting and machine tool accessory, rolling mill, and other metalworking machinery manufacturing',
- 'Turbine and turbine generator set units manufacturing',
- 'Speed changer, industrial high-speed drive, and gear manufacturing',
- 'Mechanical power transmission equipment manufacturing',
- 'Other engine equipment manufacturing',
- 'Air and gas compressor manufacturing',
- 'Pump and pumping equipment manufacturing',
- 'Material handling equipment manufacturing',
- 'Power-driven handtool manufacturing',
- 'Packaging machinery manufacturing',
- 'Industrial process furnace and oven manufacturing',
- 'Other general purpose machinery manufacturing',
- 'Fluid power process machinery',
- 'Electronic computer manufacturing',
- 'Computer storage device manufacturing',
- 'Computer terminals and other computer peripheral equipment manufacturing',
- 'Telephone apparatus manufacturing',

'Broadcast and wireless communications equipment',
 'Other communications equipment manufacturing',
 'Audio and video equipment manufacturing',
 'Semiconductor and related device manufacturing',
 'Printed circuit assembly (electronic assembly) manufacturing',
 'Other electronic component manufacturing',
 'Electromedical and electrotherapeutic apparatus manufacturing',
 'Search, detection, and navigation instruments manufacturing',
 'Automatic environmental control manufacturing',
 'Industrial process variable instruments manufacturing',
 'Totalizing fluid meter and counting device manufacturing',
 'Electricity and signal testing instruments manufacturing',
 'Analytical laboratory instrument manufacturing',
 'Irradiation apparatus manufacturing',
 'Watch, clock, and other measuring and controlling device manufacturin
 g',
 'Manufacturing and reproducing magnetic and optical media',
 'Electric lamp bulb and part manufacturing',
 'Lighting fixture manufacturing',
 'Small electrical appliance manufacturing',
 'Household cooking appliance manufacturing',
 'Household refrigerator and home freezer manufacturing',
 'Household laundry equipment manufacturing',
 'Other major household appliance manufacturing',
 'Power, distribution, and specialty transformer manufacturing',
 'Motor and generator manufacturing',
 'Switchgear and switchboard apparatus manufacturing',
 'Relay and industrial control manufacturing',
 'Storage battery manufacturing', 'Primary battery manufacturing',
 'Communication and energy wire and cable manufacturing',
 'Wiring device manufacturing',
 'Carbon and graphite product manufacturing',
 'All other miscellaneous electrical equipment and component manufactur
 ing',
 'Automobile manufacturing',
 'Light truck and utility vehicle manufacturing',
 'Heavy duty truck manufacturing',
 'Motor vehicle body manufacturing', 'Truck trailer manufacturing',
 'Motor home manufacturing',
 'Travel trailer and camper manufacturing',
 'Motor vehicle gasoline engine and engine parts manufacturing',
 'Motor vehicle electrical and electronic equipment manufacturing',
 'Motor vehicle transmission and power train parts manufacturing',
 'Motor vehicle seating and interior trim manufacturing',
 'Motor vehicle metal stamping',
 'Other Motor Vehicle Parts Manufacturing',
 'Motor vehicle steering, suspension component (except spring), and bra
 ke systems manufacturing',
 'Aircraft manufacturing',
 'Aircraft engine and engine parts manufacturing',
 'Other aircraft parts and auxiliary equipment manufacturing',
 'Guided missile and space vehicle manufacturing',
 'Railroad rolling stock manufacturing',
 'Ship building and repairing', 'Boat building',

'Motorcycle, bicycle, and parts manufacturing',
'Military armored vehicle, tank, and tank component manufacturing',
'All other transportation equipment manufacturing',
'Wood kitchen cabinet and countertop manufacturing',
'Upholstered household furniture manufacturing',
'Nonupholstered wood household furniture manufacturing',
'Institutional furniture manufacturing',
'Showcase, partition, shelving, and locker manufacturing',
'Office furniture and custom architectural woodwork and millwork manufacturing',
'Other furniture related product manufacturing',
'Surgical and medical instrument manufacturing',
'Surgical appliance and supplies manufacturing',
'Dental equipment and supplies manufacturing',
'Ophthalmic goods manufacturing',
'Jewelry and silverware manufacturing',
'Sporting and athletic goods manufacturing',
'Doll, toy, and game manufacturing',
'Office supplies (except paper) manufacturing',
'Sign manufacturing', 'All other miscellaneous manufacturing',
'Motor vehicle and parts dealers', 'Air transportation',
'Rail transportation', 'Water transportation',
'Truck transportation',
'Transit and ground passenger transportation',
'Pipeline transportation', 'Warehousing and storage',
'Newspaper publishers', 'Periodical Publishers', 'Book publishers',
'Software publishers', 'Motion picture and video industries',
'Sound recording industries', 'Radio and television broadcasting',
'Cable and other subscription programming',
'Wired telecommunications carriers',
'Wireless telecommunications carriers (except satellite)',
'Data processing, hosting, and related services',
'Internet publishing and broadcasting and Web search portals',
'News syndicates, libraries, archives and all other information services',
'Other financial investment activities',
'Securities and commodity contracts intermediation and brokerage',
'Insurance agencies, brokerages, and related activities',
'Funds, trusts, and other financial vehicles',
'Automotive equipment rental and leasing',
'Commercial and industrial machinery and equipment rental and leasing',
'General and consumer goods rental',
'Accounting, tax preparation, bookkeeping, and payroll services',
'Specialized design services',
'Custom computer programming services',
'Computer systems design services',
'Management consulting services',
'Scientific research and development services',
'Advertising, public relations, and related services',
'Photographic services', 'Veterinary services',
'All other miscellaneous professional, scientific, and technical services',
'Management of companies and enterprises',

```

'Office administrative services', 'Facilities support services',
'Business support services',
'Travel arrangement and reservation services',
'Investigation and security services',
'Services to buildings and dwellings',
'Waste management and remediation services',
'Junior colleges, colleges, universities, and professional schools',
'Offices of physicians', 'Offices of dentists',
'Offices of other health practitioners', 'Outpatient care centers',
'Medical and diagnostic laboratories', 'Home health care services',
'Other ambulatory health care services',
'Residential mental health, substance abuse, and other residential care facilities',
'Child day care services', 'Performing arts companies',
'Spectator sports',
'Promoters of performing arts and sports and agents for public figures',
'Gambling industries (except casino hotels)',
'Other amusement and recreation industries', 'Accommodation',
'Automotive repair and maintenance',
'Electronic and precision equipment repair and maintenance',
'Commercial and industrial machinery and equipment repair and maintenance',
'Personal and household goods repair and maintenance',
'Personal care services', 'Death care services',
'Dry-cleaning and laundry services', 'Other personal services',
'Private households'], dtype=object)

```

```
In [116... len(df.Name.unique()) # Checking number of unique values in 'Name'
```

```
Out[116... 713
```

Top 10 Emitting Industry

```
In [117... top_emitters = df[['Name', 'Supply Chain Emission Factors with Margins']].groupby(
    'Supply Chain Emission Factors with Margins', ascending=False).head(10)

# Resetting index for better plotting
top_emitters = top_emitters.reset_index()
```

```
In [118... top_emitters
```

Out[118...

	Name	Supply Chain Emission Factors with Margins
0	Cement manufacturing	1.686179
1	Cement	1.324964
2	Electric power generation, transmission, and d...	1.220357
3	Electricity	1.016143
4	Dolls, toys, and games	0.832179
5	Lime and gypsum products	0.816536
6	Lime and gypsum product manufacturing	0.799679
7	Industrial gas manufacturing	0.612929
8	Compressed Gases	0.539679
9	Clothing	0.468714

In [119...

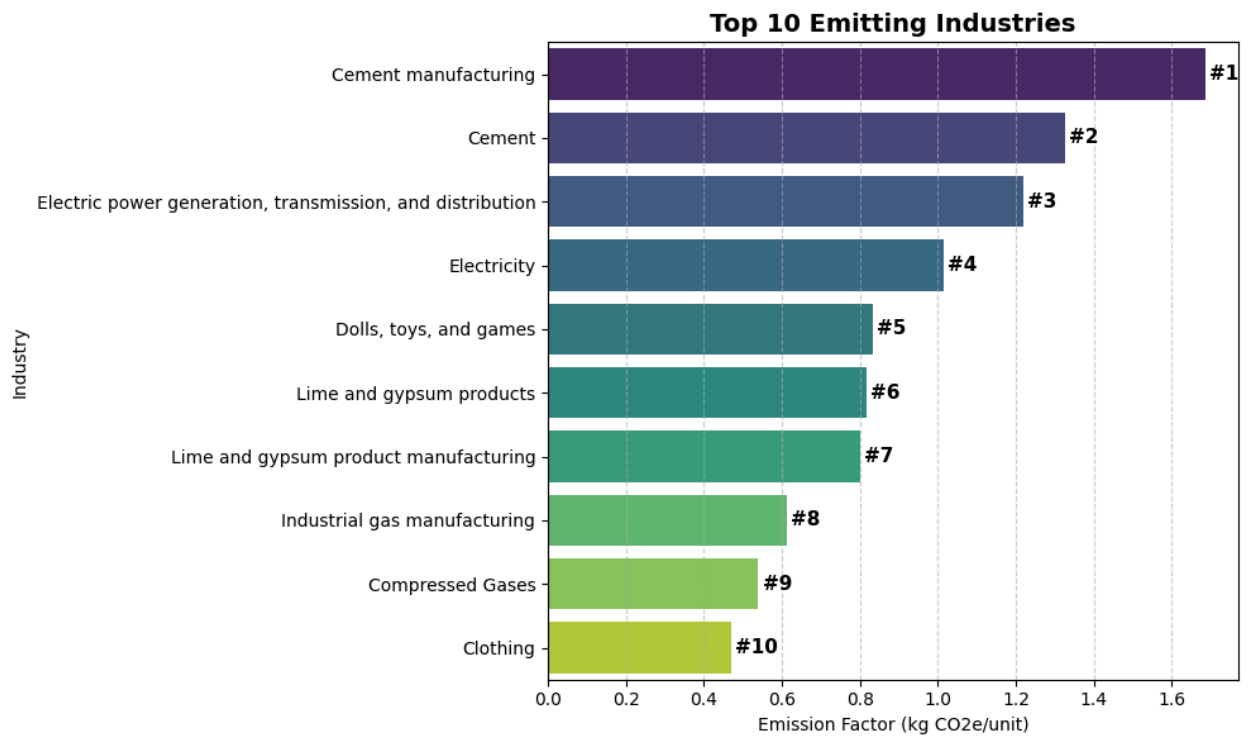
```
# Plotting the top 10 emitting industries
```

```
plt.figure(figsize=(10,6))
# Example: Top emitting industries (already grouped)
sns.barplot(
    x='Supply Chain Emission Factors with Margins',
    y='Name',
    data=top_emitters,
    hue='Name',
    palette='viridis' # Use 'Blues', 'viridis', etc., for other color maps
)

# Add ranking labels (1, 2, 3...) next to bars
for i, (value, name) in enumerate(zip(top_emitters['Supply Chain Emission Fact
    plt.text(value + 0.01, i - 1, f'#{i}', va='center', fontsize=11, fontweigh

plt.title('Top 10 Emitting Industries', fontsize=14, fontweight='bold') # Titl
plt.xlabel('Emission Factor (kg CO2e/unit)') # X-axis label
plt.ylabel('Industry') # Y-axis label
plt.grid(axis='x', linestyle='--', alpha=0.6) # Adding grid lines for better r
plt.tight_layout() # Adjust layout to prevent overlap

plt.show()
```



```
In [120...] df.drop(columns=['Name', 'Code', 'Year'], inplace=True)
```

```
In [121...] df.head(1)
```

```
Out[121...]
      Substance  Unit  Supply Chain Emission Factors without Margins  Margins of Supply Chain Emission Factors  Supply Chain Emission Factors with Margins  ReliabilityScore of Factors without Margins  DQ  TemporalCor of Factors
```

0	0	0	0.398	0.073	0.47		4	
---	---	---	-------	-------	------	--	---	--

```
In [122...] df.shape
```

```
Out[122...] (22092, 11)
```

Defining features and target

```
In [123...] df.columns
```

```
Out[123... Index(['Substance', 'Unit', 'Supply Chain Emission Factors without Margins',
      'Margins of Supply Chain Emission Factors',
      'Supply Chain Emission Factors with Margins',
      'DQ ReliabilityScore of Factors without Margins',
      'DQ TemporalCorrelation of Factors without Margins',
      'DQ GeographicalCorrelation of Factors without Margins',
      'DQ TechnologicalCorrelation of Factors without Margins',
      'DQ DataCollection of Factors without Margins', 'Source'],
      dtype='object')
```

```
In [124... X = df.drop(columns=['Supply Chain Emission Factors with Margins']) # Feature
y = df['Supply Chain Emission Factors with Margins'] # Target variable
```

```
In [125... X.head()
```

```
Out[125...
```

	Substance	Unit	Supply Chain Emission Factors without Margins	Margins of Supply Chain Emission Factors	DQ ReliabilityScore of Factors without Margins	DQ TemporalCorrelation of Factors without Margins
0	0	0	0.398	0.073	4	3
1	1	0	0.001	0.001	4	3
2	2	0	0.002	0.000	4	3
3	3	1	0.002	0.000	3	3
4	0	0	0.659	0.081	4	3

```
In [126... y.head()
```

```
Out[126...
```

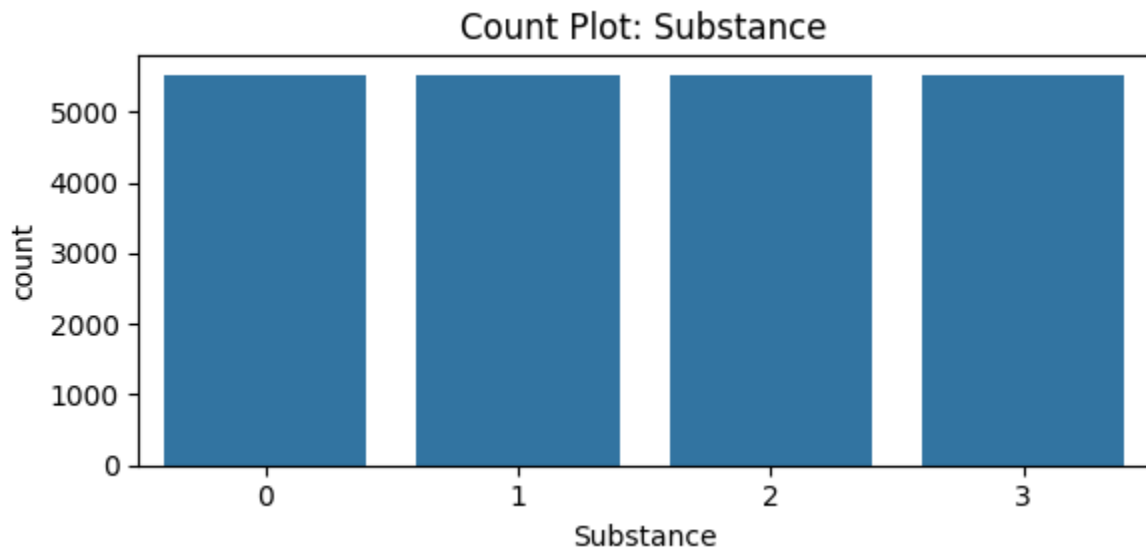
	Supply Chain Emission Factors with Margins
0	0.470
1	0.002
2	0.002
3	0.002
4	0.740

dtype: float64

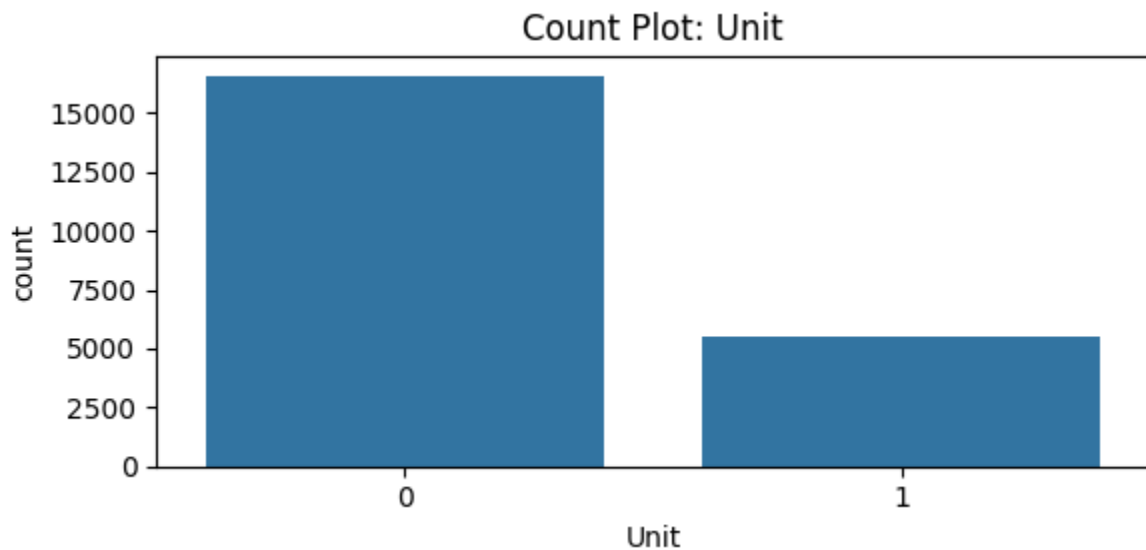
Univariate Analysis

```
In [127... # Count plot for Substance
plt.figure(figsize=(6, 3))
```

```
sns.countplot(x=df["Substance"])
plt.title("Count Plot: Substance")
plt.xticks()
plt.tight_layout()
plt.show()
```



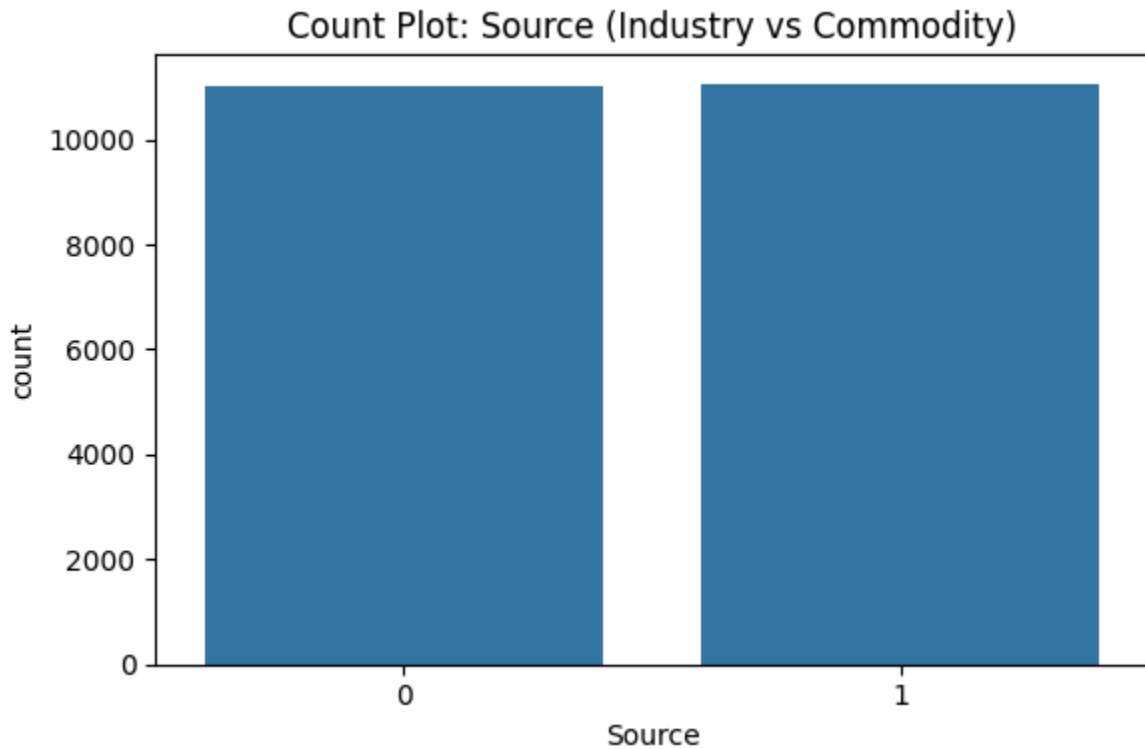
```
In [128... # Count plot for Unit
plt.figure(figsize=(6, 3))
sns.countplot(x=df["Unit"])
plt.title("Count Plot: Unit")
plt.tight_layout()
plt.show()
```



```
In [129... # Count plot for Source
plt.figure(figsize=(6, 4))
sns.countplot(x=df["Source"])
plt.title("Count Plot: Source (Industry vs Commodity)")
plt.tight_layout()
```



```
plt.show()
```



```
In [130...] df.columns
```

```
Out[130...] Index(['Substance', 'Unit', 'Supply Chain Emission Factors without Margins',  
                'Margins of Supply Chain Emission Factors',  
                'Supply Chain Emission Factors with Margins',  
                'DQ ReliabilityScore of Factors without Margins',  
                'DQ TemporalCorrelation of Factors without Margins',  
                'DQ GeographicalCorrelation of Factors without Margins',  
                'DQ TechnologicalCorrelation of Factors without Margins',  
                'DQ DataCollection of Factors without Margins', 'Source'],  
                dtype='object')
```

Multivariate Analysis

Correlation heatmap

```
In [131...] df.select_dtypes(include=np.number).corr() # Checking correlation between nume
```

Out[131...

	Substance	Unit	Supply Chain Emission Factors without Margins	Margins of Supply Chain Emission Factors	Supply Chain Emission Factors with Margins
Substance	1.000000e+00	7.745967e-01	-0.391851	-0.218400	-0.42
Unit	7.745967e-01	1.000000e+00	-0.155859	-0.094300	-0.16
Supply Chain Emission Factors without Margins	-3.918505e-01	-1.558594e-01	1.000000	0.143005	0.96
Margins of Supply Chain Emission Factors	-2.184002e-01	-9.429989e-02	0.143005	1.000000	0.40
Supply Chain Emission Factors with Margins	-4.216032e-01	-1.697410e-01	0.962971	0.404541	1.00
DQ ReliabilityScore of Factors without Margins	9.509190e-02	-2.515938e-02	-0.098000	-0.069598	-0.10
DQ TemporalCorrelation of Factors without Margins	-3.667637e-15	-3.173071e-17	0.009284	0.007953	0.01
DQ GeographicalCorrelation of Factors without Margins	NaN	NaN	NaN	NaN	
DQ TechnologicalCorrelation of Factors without Margins	1.984154e-01	2.869901e-01	0.148410	0.086335	0.16
DQ DataCollection of Factors without Margins	NaN	NaN	NaN	NaN	
Source	4.273306e-16	-1.545892e-17	0.027131	-0.067504	0.00

In [132...

```
df.info() # Checking data types and non-null counts after mapping
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22092 entries, 0 to 22091
Data columns (total 11 columns):
#   Column                                                                 Non-Null Count  Dt
type  -----
----
0   Substance                                                            22092 non-null  in
t64
1   Unit                                                                22092 non-null  in
t64
2   Supply Chain Emission Factors without Margins                    22092 non-null  fl
oat64
3   Margins of Supply Chain Emission Factors                        22092 non-null  fl
oat64
4   Supply Chain Emission Factors with Margins                      22092 non-null  fl
oat64
5   DQ ReliabilityScore of Factors without Margins                  22092 non-null  in
t64
6   DQ TemporalCorrelation of Factors without Margins              22092 non-null  in
t64
7   DQ GeographicalCorrelation of Factors without Margins          22092 non-null  in
t64
8   DQ TechnologicalCorrelation of Factors without Margins         22092 non-null  in
t64
9   DQ DataCollection of Factors without Margins                   22092 non-null  in
t64
10  Source                                                                22092 non-null  in
t64
dtypes: float64(3), int64(8)
memory usage: 1.9 MB

```

```

In [133... # Correlation matrix
plt.figure(figsize=(12, 8))
sns.heatmap(df.select_dtypes(include=np.number).corr(), annot=True, cmap="cool
plt.title("Correlation Heatmap")
plt.show()

```

