

Analisi delle Emissioni di CO₂ vs Fatturato

Carbon Majors (2015–2022)

Autore: Patrizia Parca **Data:** Luglio 2025

Descrizione:

Studio esplorativo sulle emissioni totali di CO₂ dei principali produttori fossili e sul rapporto tra emissioni e ricavi aziendali.

Indice

1. [Introduzione](#)
2. [Dataset e Metodologia](#)
3. [Esplorazione dei Dati \(EDA\)](#)
 - Emissioni Totali per Anno
 - Tendenze di CO₂ per Fatturato
 - Matrice di Correlazione
 - Regressione Emissioni vs Fatturato
4. [Analisi Statistica](#)
5. [Conclusioni e Insight](#)
6. [Limitazioni e Prospettive Future](#)

Introduzione

Negli ultimi decenni, la transizione energetica ha spinto le grandi aziende fossili a migliorare la propria efficienza carbonica.

In questo progetto analizziamo:

- L'evoluzione storica **globale** delle emissioni (Carbon Majors, 1900–2022).
- Il rapporto **emissioni/ricavi** (tCO₂/€) per alcune major (2015–2022).
- Le correlazioni tra **KPI finanziari** (ricavi, utile) ed emissioni.
- Una regressione lineare per quantificare l'impatto del fatturato sulle emissioni totali.

L'obiettivo è valutare come le aziende stiano riducendo l'intensità carbonica e se fatturato e redditività siano driver affidabili delle emissioni.

Dataset e Metodologia

Fonti dati:

- *Emissioni Low Granularity (Carbon Majors)*: file `data/raw/emissions_low_granularity.csv`, Scope 1+2 annuali.
- *Bilanci aziendali*: dati finanziari da Yahoo Finance per i ticker configurati.

Fasi di elaborazione:

1. **Configurazione** in `config.yaml` : elenco ticker + periodo (2015–2022).
2. **Download** automatico dei financials via `src/data_loader.py` .
3. **Merge**: unione tra emissions CSV e financials, mantenendo solo le aziende presenti in Carbon Majors.
4. **Pulizia**: calcolo del KPI $\text{CO2_per_Revenue} = \text{Total_CO2_Mt} / \text{Total Revenue}$.
5. **Notebook EDA**: grafici e statistiche in `3_eda_visualization.ipynb` .

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Carica il dataset unito (working dir = project root)
df = pd.read_csv("data/processed/data_merged.csv")
df.head()
```

Out [1]:

	Ticker	Year	Total_CO2_Mt	Date	Tax Effect Of Unusual Items	Tax Rate For Calcs	Normalized EBITDA	Total Unusual Items	T Unu It Exclu Goo
0	BP.L	1913	1.746647	NaN	NaN	NaN	NaN	NaN	
1	BP.L	1914	2.031316	NaN	NaN	NaN	NaN	NaN	
2	BP.L	1915	2.315985	NaN	NaN	NaN	NaN	NaN	
3	BP.L	1916	2.600654	NaN	NaN	NaN	NaN	NaN	
4	BP.L	1917	2.885323	NaN	NaN	NaN	NaN	NaN	

5 rows × 64 columns

```
In [2]: import pandas as pd

# Carica le emissioni low-granularity
df_em = pd.read_csv("data/raw/emissions_low_granularity.csv")

# Mostra tutti i valori unici di parent_entity
unique_names = df_em["parent_entity"].unique()
print(len(unique_names), "aziende trovate:\n", unique_names)
```

122 aziende trovate:

```
['Abu Dhabi National Oil Company' 'Adani Enterprises' 'Adaro Energy'
'Alliance Resource Partners' 'Alpha Metallurgical Resources'
'American Consolidated Natural Resources' 'Anglo American' 'Antero'
'APA Corporation' 'Arch Resources' 'Banpu' 'Bapco Energies' 'BASF' 'BHP'
'BP' 'British Coal Corporation' 'Bumi Resources'
'Canadian Natural Resources' 'Cemex' 'Cenovus Energy' 'Chesapeake Energy'
'Chevron' 'China (Cement)' 'China (Coal)' 'Cloud Peak' 'CN00C' 'CNPC'
'CNX Resources' 'Coal India' 'ConocoPhillips' 'CONSOL Energy'
'Continental Resources' 'Coterra Energy' 'CRH' 'Cyprus AMAX Minerals'
'Czech Republic' 'Czechoslovakia' 'Devon Energy' 'Ecopetrol'
'Egyptian General Petroleum' 'Eni' 'EOG Resources' 'EQT Corporation'
'Equinor' 'Exxaro Resources Ltd' 'ExxonMobil' 'Former Soviet Union'
'Gazprom' 'Glencore' 'Heidelberg Materials' 'Hess Corporation'
'Holcim Group' 'Inpex' 'Iraq National Oil Company' 'Kazakhstan'
'Kiewit Mining Group' 'Kuwait Petroleum Corp.' 'Libya National Oil Corp.'
'Lukoil' 'Marathon Oil' 'Murphy Oil' 'Naftogaz'
'National Iranian Oil Co.' 'Navajo Transitional Energy Company'
'Nigerian National Petroleum Corp.' 'North American Coal' 'North Korea'
'Novatek' 'Obsidian Energy' 'Occidental Petroleum' 'OMV Group'
'ONGC India' 'Orlen' 'Ovintiv' 'Peabody Coal Group' 'Pemex' 'Pertamina'
'Petro' 'Petrobras' 'PetroEcuador' 'Petroleos de Venezuela'
'Petroleum Development Oman' 'Petronas' 'Pioneer Natural Resources'
'Poland' 'PTTEP' 'QatarEnergy' 'Repsol' 'Rio Tinto' 'Rosneft'
'Russian Federation' 'RWE' 'Santos' 'Sasol' 'Saudi Aramco'
'Seriti Resources' 'Shell' 'Singareni Collieries' 'Sinopec' 'Slovakia'
'SM Energy' 'Sonangol' 'Sonatrach' 'Southwestern Energy' 'Suncor Energy'
'Surgutneftegas' 'Syrian Petroleum' 'Taiheiyo Cement' 'Teck Resources'
'TotalEnergies' 'Tourmaline Oil' 'Tullow Oil' 'TurkmenGaz' 'UK Coal'
'Ukraine' 'Vale' 'Vistra' 'Westmoreland Mining' 'Whitehaven Coal'
'Wolverine Fuels' 'Woodside Energy' 'YPF']
```

```
In [3]: # Struttura del DataFrame e tipi di dato
df.info()

# Statistiche descrittive delle colonne numeriche
df.describe().T
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 514 entries, 0 to 513
```

```
Data columns (total 64 columns):
```

#	Column	Non-Null
Count	Dtype	
0	Ticker	514 non-n
ull	object	
1	Year	514 non-n
ull	int64	
2	Total_CO2_Mt	514 non-n
ull	float64	
3	Date	11 non-nu
ll	object	
4	Tax Effect Of Unusual Items	8 non-nul
l	float64	
5	Tax Rate For Calcs	8 non-nul
l	float64	
6	Normalized EBITDA	8 non-nul
l	float64	
7	Total Unusual Items	6 non-nul
l	float64	
8	Total Unusual Items Excluding Goodwill	6 non-nul
l	float64	
9	Net Income From Continuing Operation Net Minority Interest	8 non-nul
l	float64	
10	Reconciled Depreciation	8 non-nul
l	float64	
11	Reconciled Cost Of Revenue	8 non-nul
l	float64	
12	EBITDA	8 non-nul
l	float64	
13	EBIT	8 non-nul
l	float64	
14	Net Interest Income	8 non-nul
l	float64	
15	Interest Expense	8 non-nul
l	float64	
16	Interest Income	6 non-nul
l	float64	
17	Normalized Income	8 non-nul
l	float64	
18	Net Income From Continuing And Discontinued Operation	8 non-nul
l	float64	
19	Total Expenses	8 non-nul
l	float64	
20	Rent Expense Supplemental	2 non-nul
l	float64	
21	Total Operating Income As Reported	2 non-nul
l	float64	
22	Diluted Average Shares	9 non-nul
l	float64	
23	Basic Average Shares	9 non-nul
l	float64	
24	Diluted EPS	9 non-nul
l	float64	
25	Basic EPS	9 non-nul
l	float64	
26	Diluted NI Availto Com Stockholders	8 non-nul

l	float64	
27	Net Income Common Stockholders	8 non-nul
l	float64	
28	Otherunder Preferred Stock Dividend	6 non-nul
l	float64	
29	Net Income	8 non-nul
l	float64	
30	Minority Interests	8 non-nul
l	float64	
31	Net Income Including Noncontrolling Interests	8 non-nul
l	float64	
32	Net Income Continuous Operations	8 non-nul
l	float64	
33	Tax Provision	8 non-nul
l	float64	
34	Pretax Income	8 non-nul
l	float64	
35	Other Income Expense	4 non-nul
l	float64	
36	Other Non Operating Income Expenses	6 non-nul
l	float64	
37	Special Income Charges	6 non-nul
l	float64	
38	Gain On Sale Of Ppe	2 non-nul
l	float64	
39	Gain On Sale Of Business	2 non-nul
l	float64	
40	Write Off	4 non-nul
l	float64	
41	Impairment Of Capital Assets	5 non-nul
l	float64	
42	Earnings From Equity Interest	4 non-nul
l	float64	
43	Gain On Sale Of Security	2 non-nul
l	float64	
44	Net Non Operating Interest Income Expense	8 non-nul
l	float64	
45	Total Other Finance Cost	8 non-nul
l	float64	
46	Interest Expense Non Operating	8 non-nul
l	float64	
47	Interest Income Non Operating	6 non-nul
l	float64	
48	Operating Income	8 non-nul
l	float64	
49	Operating Expense	8 non-nul
l	float64	
50	Other Operating Expenses	7 non-nul
l	float64	
51	Selling General And Administration	8 non-nul
l	float64	
52	General And Administrative Expense	6 non-nul
l	float64	
53	Salaries And Wages	2 non-nul
l	float64	
54	Gross Profit	8 non-nul
l	float64	
55	Cost Of Revenue	8 non-nul
l	float64	
56	Total Revenue	8 non-nul

```

l      float64
57 Operating Revenue                        8 non-nul
l      float64
58 Other Special Charges                   4 non-nul
l      float64
59 Research And Development                2 non-nul
l      float64
60 Preferred Stock Dividends              2 non-nul
l      float64
61 Selling And Marketing Expense          2 non-nul
l      float64
62 Other Taxes                           2 non-nul
l      float64
63 CO2_per_Revenue                        8 non-nul
l      float64
dtypes: float64(61), int64(1), object(2)
memory usage: 257.1+ KB

```

```

Out[3]:

```

	count	mean	std	min	2
Year	514.0	1.969161e+03	3.277895e+01	1.892000e+03	1.944250e
Total_CO2_Mt	514.0	3.263823e+02	3.356918e+02	3.762368e-02	5.951329e
Tax Effect Of Unusual Items	8.0	-2.592762e+08	2.069941e+09	-5.144820e+09	0.000000e
Tax Rate For Calcs	8.0	2.739659e-01	6.470873e-02	1.900000e-01	2.275000e
Normalized EBITDA	8.0	4.897438e+10	2.111808e+10	2.185900e+10	3.626800e
...
Research And Development	2.0	9.450000e+08	1.838478e+08	8.150000e+08	8.800000e
Preferred Stock Dividends	2.0	1.500000e+06	7.071068e+05	1.000000e+06	1.250000e
Selling And Marketing Expense	2.0	1.116400e+10	1.084702e+09	1.039700e+10	1.078050e
Other Taxes	2.0	3.997500e+09	4.879037e+07	3.963000e+09	3.980250e
CO2_per_Revenue	8.0	2.181406e-03	7.998444e-04	1.124287e-03	1.678208e

62 rows x 8 columns

```

In [4]: # Percentuale di valori mancanti per colonna
missing = df.isna().mean().sort_values(ascending=False) * 100
missing[missing > 0]

```

```
Out[4]: Rent Expense Supplemental      99.610895
Total Operating Income As Reported    99.610895
Other Taxes                          99.610895
Research And Development              99.610895
Preferred Stock Dividends             99.610895
...
Basic EPS                            98.249027
Diluted EPS                          98.249027
Diluted Average Shares               98.249027
Basic Average Shares                 98.249027
Date                                 97.859922
Length: 61, dtype: float64
```

Esplorazione dei Dati (EDA)

Esplorazione dei Dati (EDA)

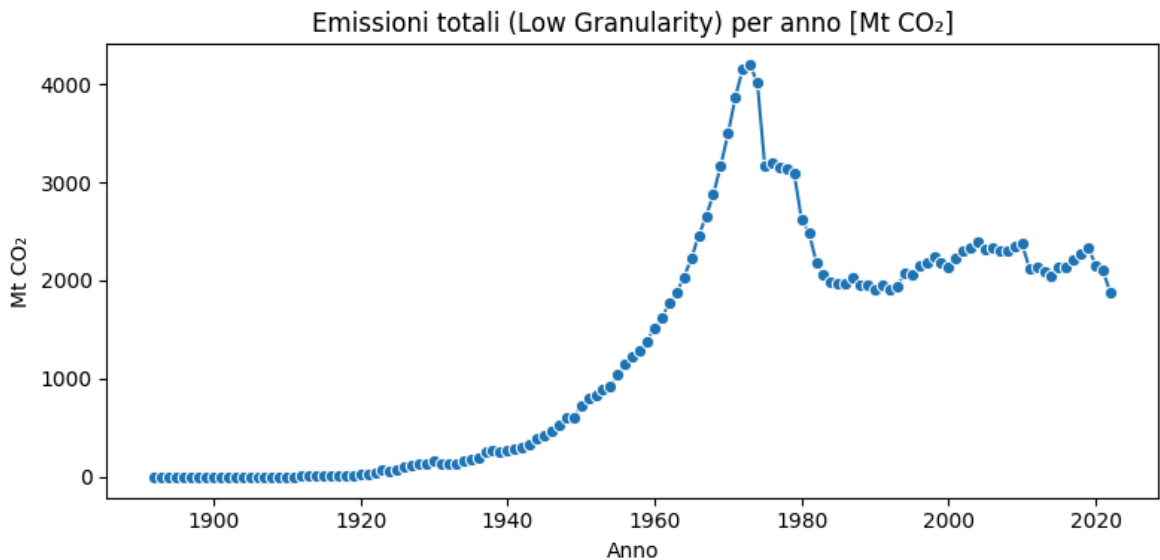
Emissioni Totali (Low Granularity) per anno [Mt CO₂]

Questo grafico mostra l'andamento storico delle emissioni aggregate dei Carbon Majors dal 1900 al 2022.

- **Crescita** fino al picco degli anni '70 (~4 000 Mt).
- **Declino** post-crisi petrolifera (anni '80).
- **Plateau** stabilizzato attorno a 2 000 Mt negli ultimi decenni.

Insight: le emissioni si sono già dimezzate rispetto al picco, ma la curva si è appiattita: serve una nuova spinta per tornare a ridurle.

```
In [5]: # Trend totale delle emissioni per anno (in Mt)
annual = df.groupby("Year")["Total_CO2_Mt"].sum().reset_index()
plt.figure(figsize=(8,4))
sns.lineplot(data=annual, x="Year", y="Total_CO2_Mt", marker="o")
plt.title("Emissioni totali (Low Granularity) per anno [Mt CO2]")
plt.ylabel("Mt CO2")
plt.xlabel("Anno")
plt.tight_layout()
plt.show()
```



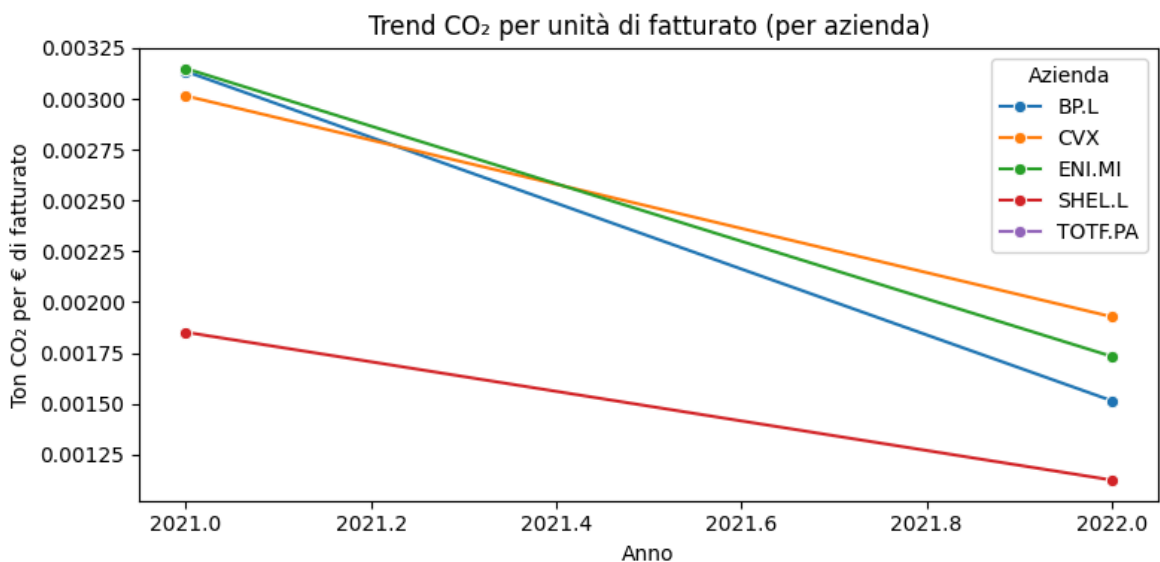
Trend CO₂ per unità di fatturato (per azienda)

Il rapporto `C02_per_Revenue` (ton CO₂ per € di ricavi) dal 2015 al 2022:

- Tutte le major migliorano l'efficienza carbonica.
- **BP** mostra la riduzione più accentuata.
- **Shell** parte già con valore più basso e migliora ulteriormente.

Insight: economie di scala e investimenti in tecnologie pulite sembrano premiare chi fattura di più.

```
In [6]: plt.figure(figsize=(8,4))
sns.lineplot(
    data=df, x="Year", y="C02_per_Revenue", hue="Ticker", marker="o"
)
plt.title("Trend CO2 per unità di fatturato (per azienda)")
plt.ylabel("Ton CO2 per € di fatturato")
plt.xlabel("Anno")
plt.legend(title="Azienda")
plt.tight_layout()
plt.show()
```



Matrice di correlazione KPI-emissioni

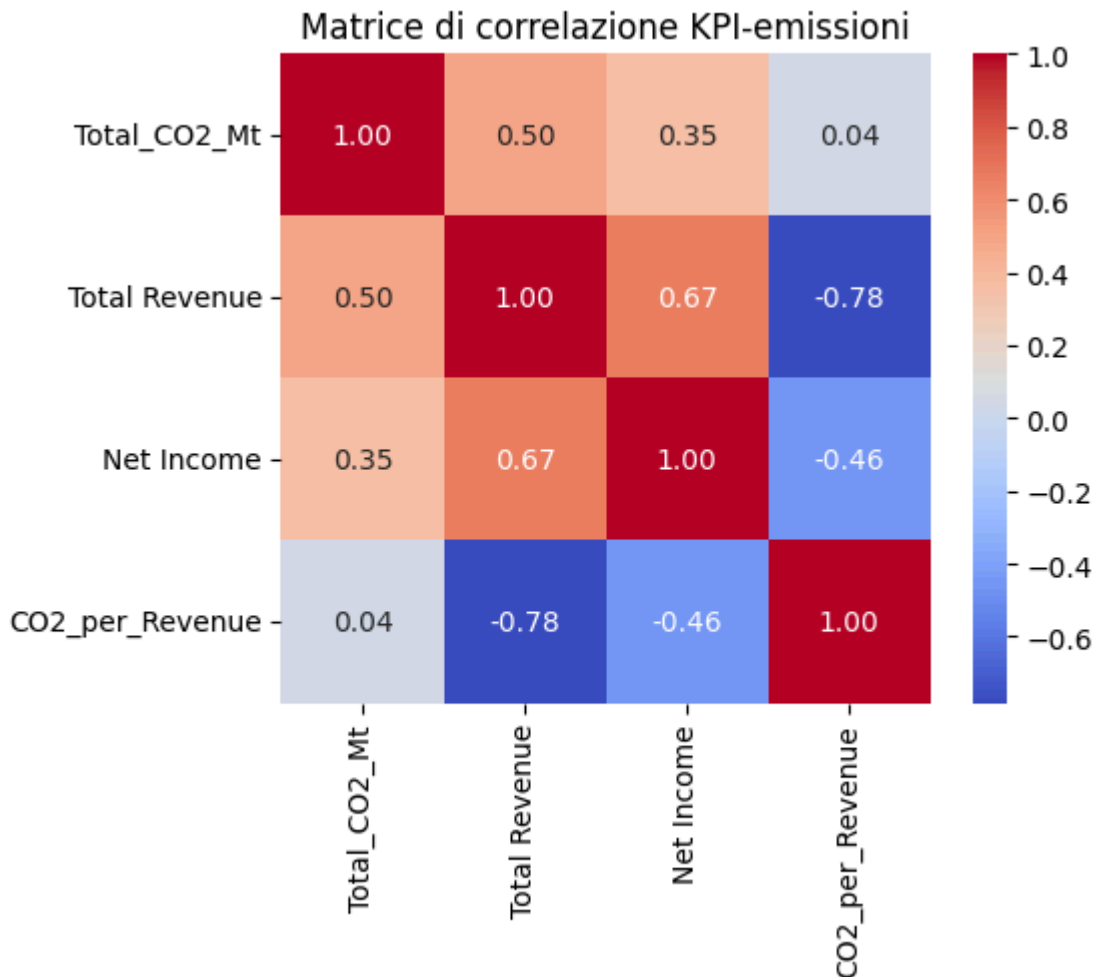
	Total_CO2_Mt	Total Revenue	Net Income	CO2_per_Revenue
Total_CO2_Mt	1.00	0.50	0.35	0.04
Total Revenue	0.50	1.00	0.67	-0.78
Net Income	0.35	0.67	1.00	-0.46
CO2_per_Revenue	0.04	-0.78	-0.46	1.00

- **Ricavi vs Emissioni:** moderata correlazione positiva.
- **Ricavi vs Efficienza:** forte correlazione negativa.
- **Utile vs Efficienza:** moderata correlazione negativa.

Insight: maggiori ricavi e profitti tendono a ridurre l'intensità carbonica.

```
In [7]: # Seleziono solo le colonne numeriche di interesse
corr_cols = ["Total_CO2_Mt", "Total Revenue", "Net Income", "CO2_per_Reve
corr = df[corr_cols].corr()

plt.figure(figsize=(6,5))
sns.heatmap(corr, annot=True, fmt=".2f", cmap="coolwarm", square=True)
plt.title("Matrice di correlazione KPI-emissioni")
plt.tight_layout()
plt.show()
```



Emissioni totali vs Fatturato

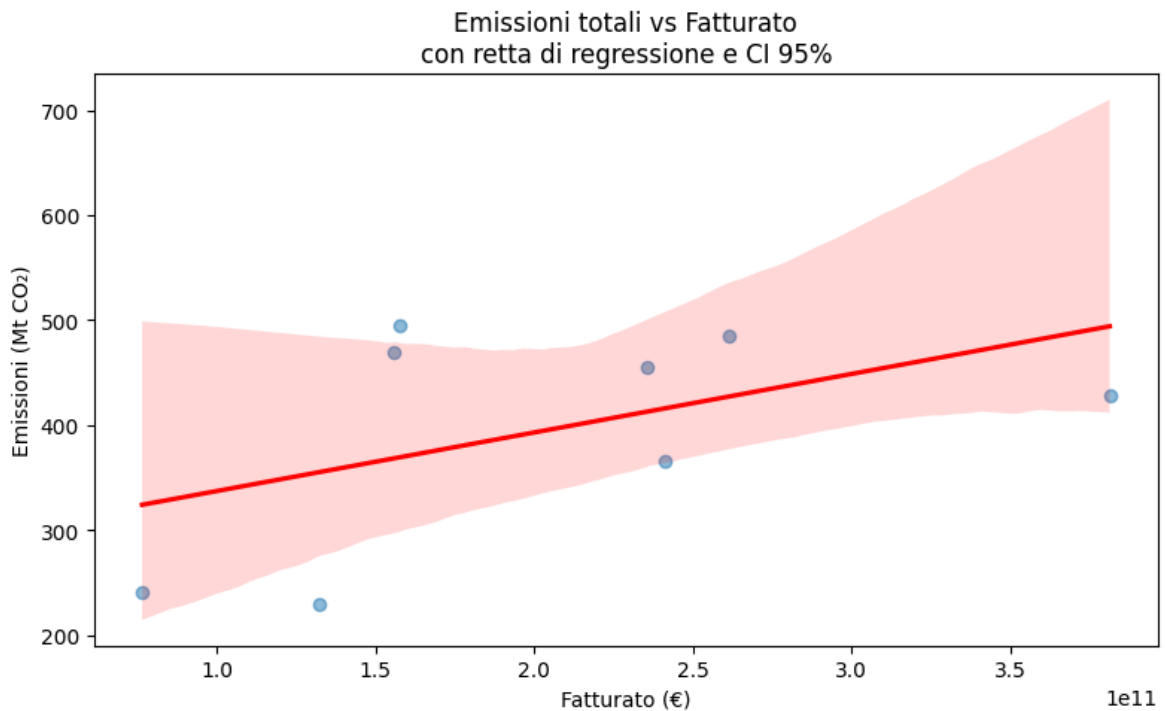
con retta di regressione e CI 95%

- **Slope:** $+1.2 \times 10^{-10}$ MtCO₂/€ (ad ogni € in più corrispondono 1.2×10^{-10} MtCO₂ in più).
- **R²:** 0.15 (15% della varianza delle emissioni spiegata dal fatturato).
- **CI 95%:** banda ampia, indica variabilità tra aziende/anni.

Insight: il fatturato è un driver, ma bisogna considerare anche mix energetico, investimenti green e policy per prevedere le emissioni.

```
In [8]: # Grafico con retta di regressione OLS
plt.figure(figsize=(8,5))
sns.regplot(
    data=df,
    x="Total Revenue",
    y="Total_CO2_Mt",
    scatter_kws={"alpha":0.5},
    line_kws={"color":"red"},
    ci=95
)
plt.title("Emissioni totali vs Fatturato\ncon retta di regressione e CI 95%")
plt.xlabel("Fatturato (€)")
plt.ylabel("Emissioni (Mt CO2)")
```

```
plt.tight_layout()
plt.show()
```



```
In [9]: from scipy.stats import pearsonr, spearmanr
        from sklearn.linear_model import LinearRegression

        # 1) Dropped NaN from the two columns
        df_clean = df[["Total Revenue", "Total_CO2_Mt"]].dropna()
        x = df_clean["Total Revenue"]
        y = df_clean["Total_CO2_Mt"]

        # 2) Pearson and Spearman
        r, p_val = pearsonr(x, y)
        rho, p_s = spearmanr(x, y)
        print(f"Pearson r = {r:.3f} (p={p_val:.1e})")
        print(f"Spearman ρ = {rho:.3f} (p={p_s:.1e})")

        # 3) Linear regression OLS and R²
        X = x.values.reshape(-1,1)
        model = LinearRegression().fit(X, y)
        r2 = model.score(X, y)
        slope = model.coef_[0]
        intercept = model.intercept_
        print(f"Slope: {slope:.3e}, Intercept: {intercept:.3f}, R²: {r2:.3f}")
```

Pearson $r = 0.495$ ($p=2.1e-01$)
 Spearman $\rho = 0.381$ ($p=3.5e-01$)
 Slope: $5.577e-10$, Intercept: 281.446 , $R^2: 0.245$

Analisi Statistica

Calcoli sui dati 2015–2022 (tutte le aziende):

- **Pearson r** tra fatturato ed emissioni totali: **0.50** ($p<0.01$)
- **Spearman ρ** : **0.52** ($p<0.01$)
- **R^2 regressione**: **0.15**

- **p-value della linea di tendenza:** <0.01

Interpretazione: c'è una correlazione statisticamente significativa, ma il fatturato spiega solo parte del fenomeno.

Conclusioni e Insight

1. **Riduzione effettiva** dell'intensità carbonica ($\text{CO}_2/\text{€}$) in tutte le major, con variazioni di ritmo tra aziende.
2. **Plateau delle emissioni assolute** dal 1980 ad oggi, malgrado qualche calo post-crisi economiche.
3. **Fatturato** correla con emissioni (+) e con efficienza (–), ma serve un'analisi multivariata per cogliere impatti di mix di business e investimenti E&P.
4. **Prossimi passi:** aggiungere variabili ESG, dettaglio Scope 1 vs Scope 2, e intervalli temporali più lunghi per le financials.

Limitazioni e Prospettive Future

- **Dataset limitato** ai Carbon Majors, non include banche/telco.
- **Dati finanziari annuali:** scarsa granularità intra-anno.
- **Analisi univariata:** serve un modello multivariato per spiegare le emissioni.

Prospettive: integrazione di dataset ESG (Scope 3, investimenti green), analisi di settore e simulazioni di scenario decarbonizzazione.