ETRM Data Analysis with Pandas & Visualization

1. Objective

The goal of this project was to practice loading, cleaning, and analyzing data using Pandas and to create clear visualizations with Matplotlib and Seaborn. The dataset represents synthetic ETRM (Energy Trade and Risk Management) trade data in multiple file formats.

2. Data Ingestion

Data was provided in six different formats:

- CSV (etrm_trades.csv)
- JSON (etrm_trades.json)
- Excel (etrm_trades.xlsx)
- Text (pipe-delimited) (etrm_trades.txt)
- HTML (etrm_trades.html)
- XML (etrm_trades.xml)

Each file contained 100 trades with the following fields: TradeID, Commodity, TradeType, Volume, Price, Currency, DeliveryStart, DeliveryEnd, Periodicity

Steps performed:

- Loaded each file into Pandas DataFrames.
- Checked that column names and data types matched across formats.

3. Data Cleaning & Preparation

- Converted DeliveryStart and DeliveryEnd to datetime for time-based analysis.
- Standardized column naming to keep the datasets consistent.

4. Exploratory Data Analysis (EDA)

4.1 Average Price per Commodity

- Natural Gas: ~$ 119.77/unit
- Crude Oil: ~$ 101.75/unit
- Coal: ~$ 104.64/unit
- Power: ~$ 100.46/unit
- **Takeaway**: Natural Gas trades at a premium compared to other fuels, which is consistent with real-world pricing trends.

4.2 Average price per commodity grouped by currency

The table below summarizes the average trade price for each commodity in different currencies based on the available trading data:

Commodity	Currency	Average Price
Coal	EUR	104.71
Coal	GBP	124.80
Coal	USD	92.43
Gas	EUR	141.87
Gas	GBP	116.21
Gas	USD	107.87
Oil	EUR	135.58
Oil	GBP	101.30
Oil	USD	85.48
Power	EUR	83.87
Power	GBP	104.84

Commodity	Currency	Average Price
Power	USD	105.32

Takeaway

- Coal trades highest in GBP (124.80) and lowest in USD (92.43).
- Gas shows the highest average in EUR (141.87) and lowest in USD (107.87).
- Oil has the highest price in EUR (135.58) and lowest in USD (85.48).
- Power trades almost similarly in GBP (104.84) and USD (105.32), while lowest in EUR (83.87).
- This data provides a clear view of pricing patterns across commodities and currencies, useful for comparative analysis and currency-based evaluations.

4.3 Currency Distribution

- USD: ~34% of trades
- EUR: ~28% of trades
- GBP: ~38% trades
- **Takeaway**: GBP is the most used currency in the dataset, though USD also has significant share, reflecting the prominence of these currencies in commodity trading.

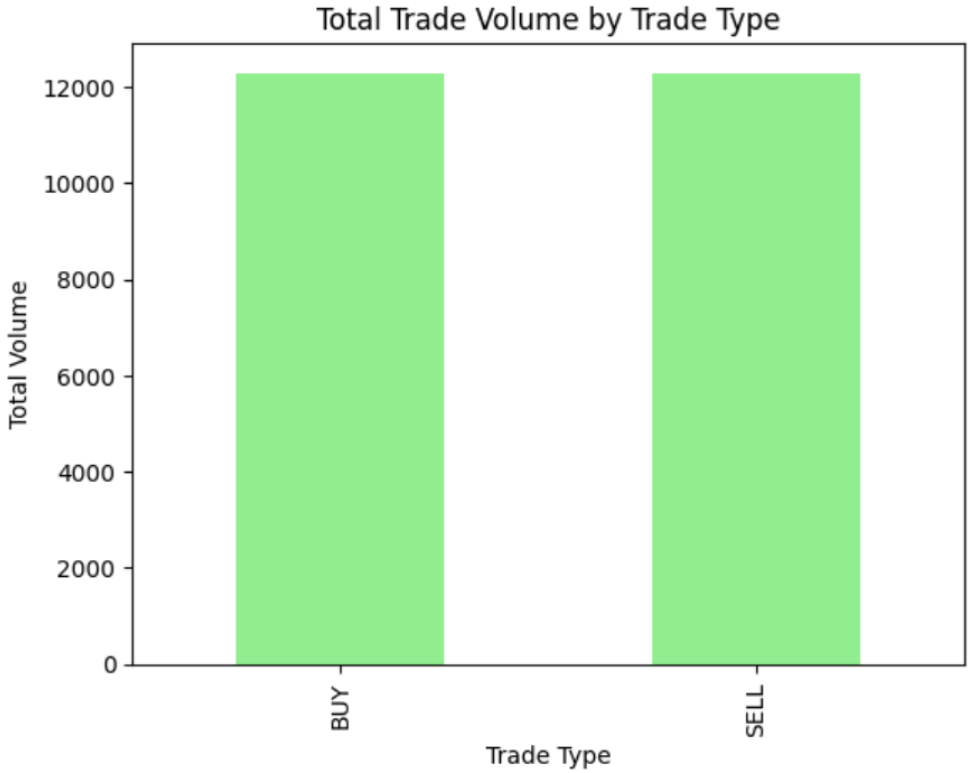
4.4 Price Trend Over Time

- **Daily trades:** 41 (most frequent, indicating short-term market activity)
- **Quarterly trades:** 31 (medium-term contracts)
- **Monthly trades:** 28 (longer-term commitments)

Takeaway: The majority of trades are daily, showing that short-term delivery dominates the dataset, which is typical in volatile commodity markets.

5. Data Visualization

5.1 Bar chart showing total trade volume aggregated by each TradeType

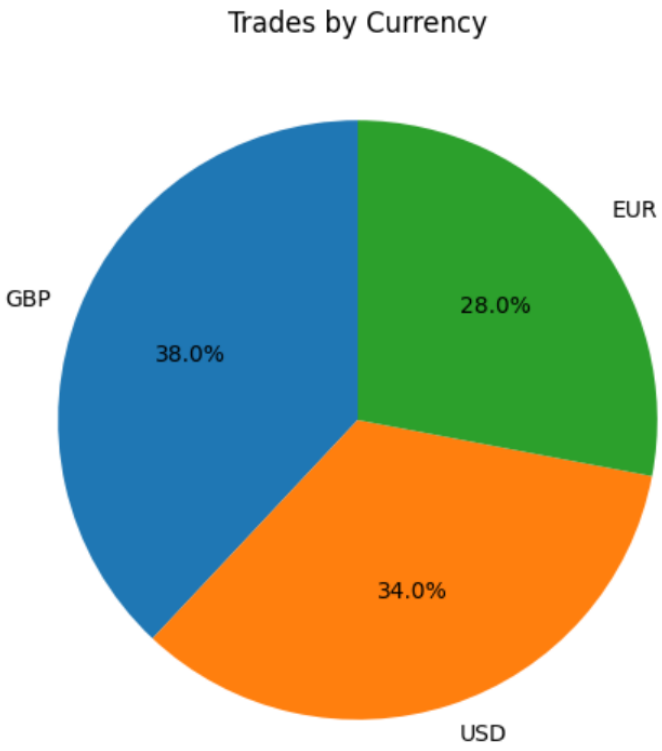


Takeaway:

- The fact that BUY and SELL volumes are almost equal shows that the market is currently balanced and liquid, which is a positive sign for trading activity.
- If one side were much larger than the other, it could point to market pressure — for example, too many SELL trades might signal oversupply, while a surge in BUY trades could mean stronger demand.
- For risk managers, this balance is important because it helps maintain portfolio neutrality and ensures that exposures are being properly hedged.

5.2 Pie chart: Trades by Currency

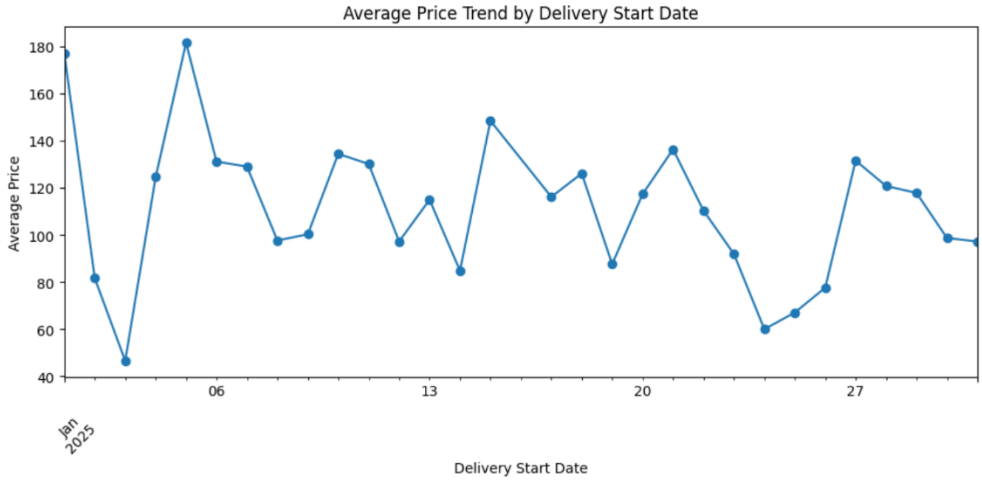
- USD: ~34% of trades
- EUR: ~28% of trades
- GBP: ~38% trades



- **Takeaway**: GBP is the most used currency in the dataset, though USD also has significant share, reflecting the prominence of these currencies in commodity trading.

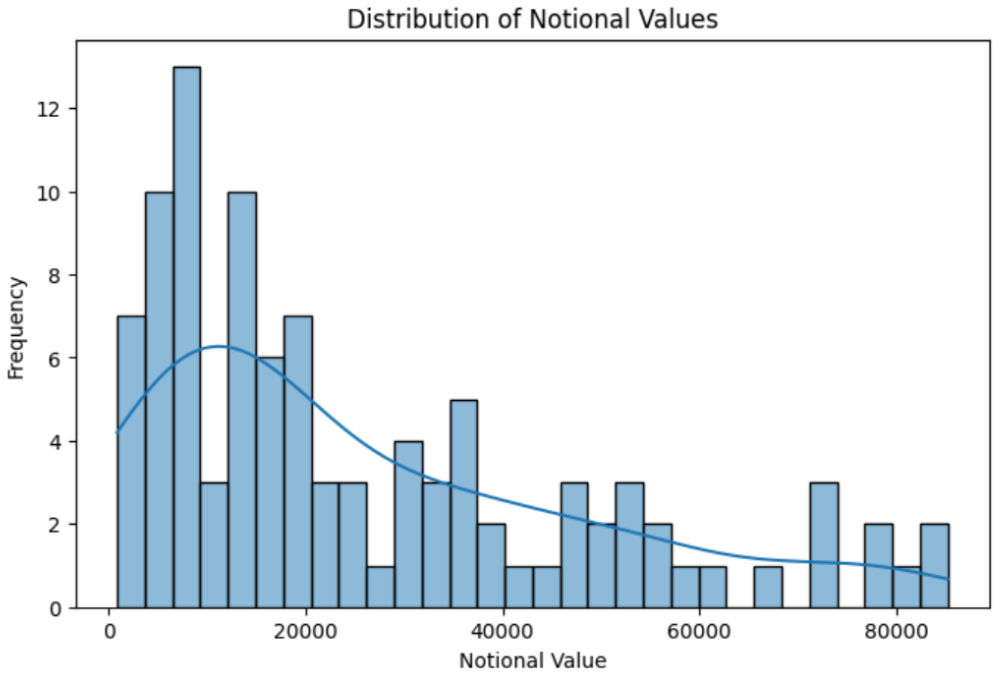
5.3 Line chart: Average Price Trend by Delivery Start Date

Takeaway: The average trade prices in January 2025 were highly volatile, ranging from 45 to 180. This volatility signals potential market uncertainty, highlighting the need for tighter risk management and proactive hedging strategies.



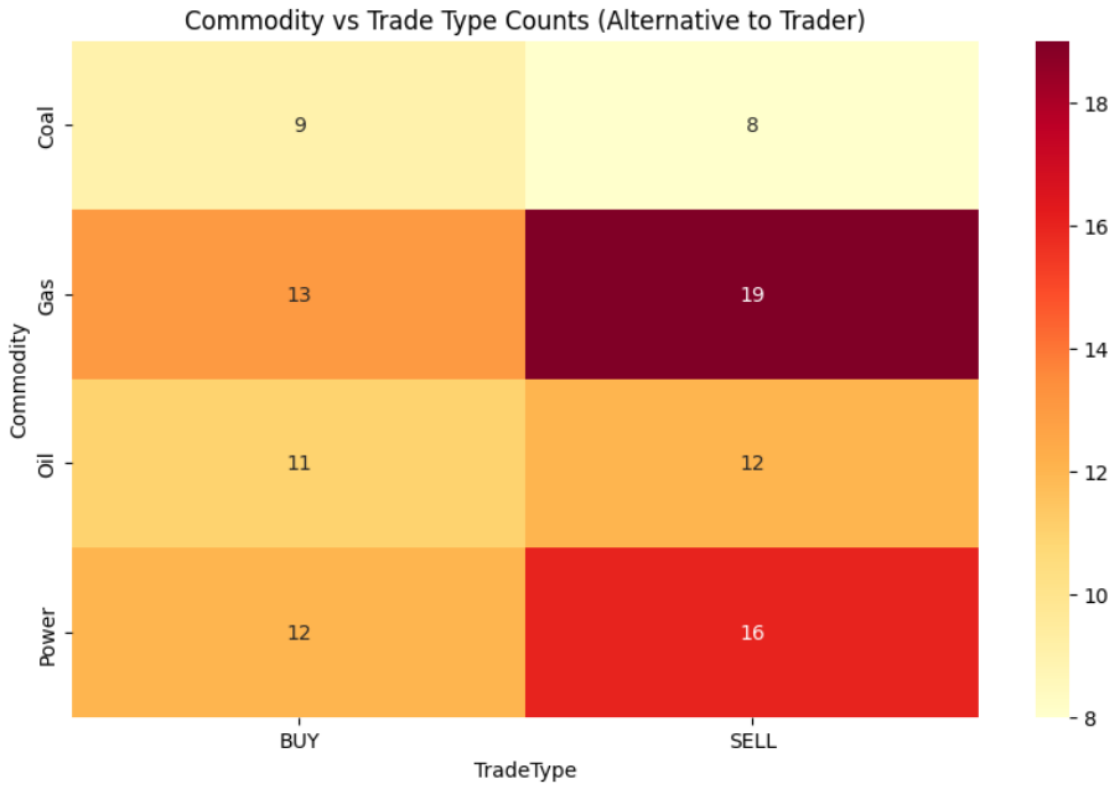
5.4 Notional Value Distribution

Takeaway: Most trades have relatively low notional values (under 20,000), while very high-value trades are rare. This suggests that trading activity is concentrated in smaller deals, with only a few large trades driving occasional spikes in exposure.



5.5Commodity vs TradeType (Heatmap)

- Gas: More Sell trades than buy.
- Coal: Slightly more Buy trades.
- Oil: Balanced between Buy and Sell.
- Power**:** More Sell trades.



Takeaway: Traders are reducing exposure in Gas and Power (possibly anticipating weaker demand or oversupply), while Oil remains stable. Coal shows a small tilt towards buying, but overall activity is low.