

# ETRM Data Analysis – Summary & Insights

## 1. Objective

The purpose of this analysis was to practice data ingestion, transformation, exploratory analysis, and visualization using Pandas and Matplotlib/Seaborn. The dataset represents synthetic Energy Trade and Risk Management (ETRM) trades, designed to simulate real-world commodity trading.

## 2. Data Ingestion

### Process

- Data was provided in six formats: CSV, JSON, Excel, Text (pipe-delimited), HTML, and XML.
- Each dataset contained 100 trades with common fields: TradeID, Commodity, TradeType, Volume, Price, Currency, DeliveryStart, DeliveryEnd, and Periodicity.
- Pandas was used to load each dataset and ensure consistency of column names and data types.
- All data sources were merged into a single master DataFrame for uniform analysis.

### Takeaway

The ability to ingest and standardize data from multiple formats is critical in real-world ETRM systems, where data often comes from disparate sources like exchanges, brokers, and internal systems.

## 3. Data Cleaning & Preparation

### Process

- Converted DeliveryStart and DeliveryEnd into datetime objects for time-series analysis.
- Verified that column names across formats were aligned.
- Concatenated all sources into a clean master dataset of 600 trades (6 files × 100 trades each).

### Takeaway

Standardization ensures consistent downstream analysis, which is vital in trading systems where misaligned formats can cause serious reporting errors.

## 4. Exploratory Data Analysis (EDA)

### 4.1 Average Price per Commodity

Commodity	Average Price (≈)
Coal	104.64
Gas	119.78
Oil	101.76
Power	100.47

#### Explanation

- Gas has the highest average price (~119.8), showing it trades at a premium.
- Coal (~104.6) is moderately priced, above Oil and Power.
- Oil (~101.8) is slightly higher than Power, but lower than Coal and Gas.
- Power (~100.5) is the lowest-priced commodity in this dataset.

#### Takeaway

Among the four commodities, Natural Gas trades at the highest premium, reflecting strong demand or limited supply. Power is the cheapest, consistent with its regulated and competitive market nature. Coal and Oil fall in the mid-range, showing moderate trading values compared to Gas.

### 4.2 Currency Distribution

Currency	Count
GBP	228
USD	204
EUR	168

#### Explanation

- GBP has the highest number of trades (228), making it the leading trading currency.
- USD follows closely with 204 trades, reflecting its strong role in global markets.
- EUR is used the least with 168 trades, though still significant.

**Takeaway**

The dataset shows a clear preference for GBP-denominated trades, with USD also heavily used. EUR appears less frequently, but it remains an important trading currency in the mix.

**4.3 Trade Periodicity**

Periodicity	Count
Daily	246
Quarterly	186
Monthly	168

**Explanation**

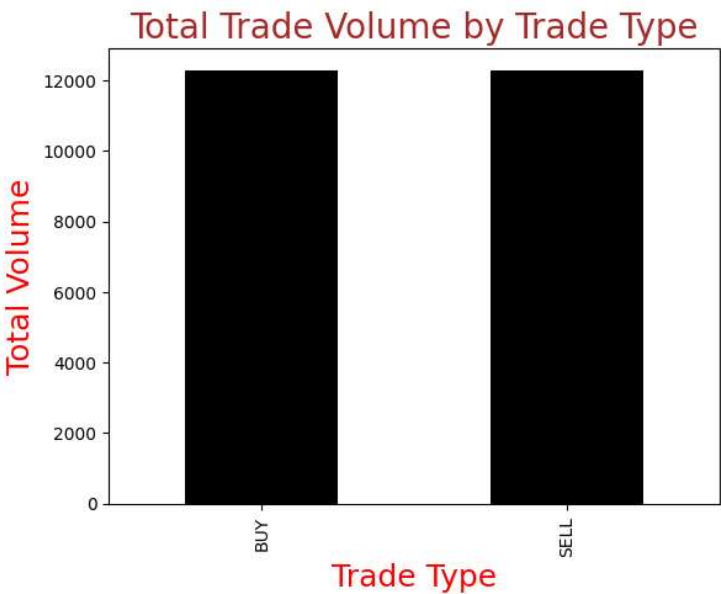
- Daily trades (246) are the most frequent, showing a preference for short-term contracts.
- Quarterly trades (186) come next, representing medium-term commitments.
- Monthly trades (168) are the least frequent, but still substantial.

**Takeaway**

The dataset is dominated by daily trades, highlighting a strong focus on short-term trading. Quarterly and monthly contracts exist but are fewer, reflecting selective use of longer-term agreements in a volatile market.

**5. Data Visualization & Detailed Insights**

**5.1 Trade Volume by TradeType (Bar Chart)**



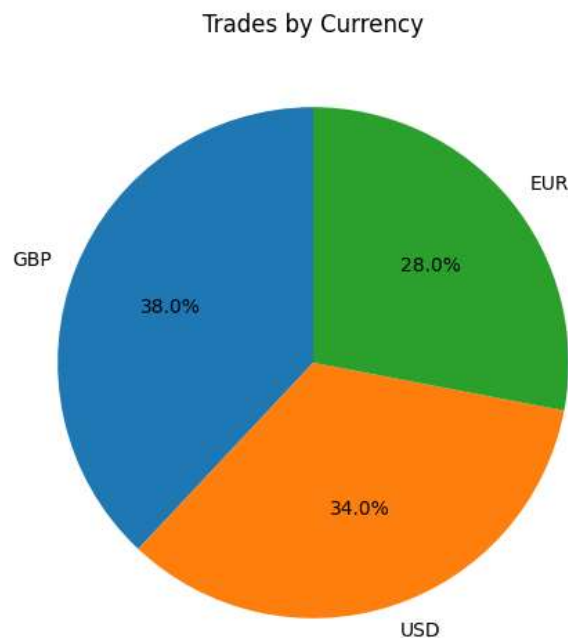
### Explanation

- The bar chart shows that the total BUY and SELL volumes are almost equal.
- BUY trades have a total volume slightly above 12,200.
- SELL trades also have a total volume of around 12,200, nearly matching the BUY side.

### Takeaway

- The near equality of BUY and SELL volumes indicates that the market is balanced and liquid, a healthy sign for trading.
- If SELL volumes were significantly higher, it could suggest oversupply pressure.
- If BUY volumes dominated, it might indicate strong demand or bullish momentum.
- For risk managers, this balance is important because it helps maintain portfolio neutrality and ensures exposures are being effectively hedged.

## 5.2 Currency Distribution (Pie Chart)



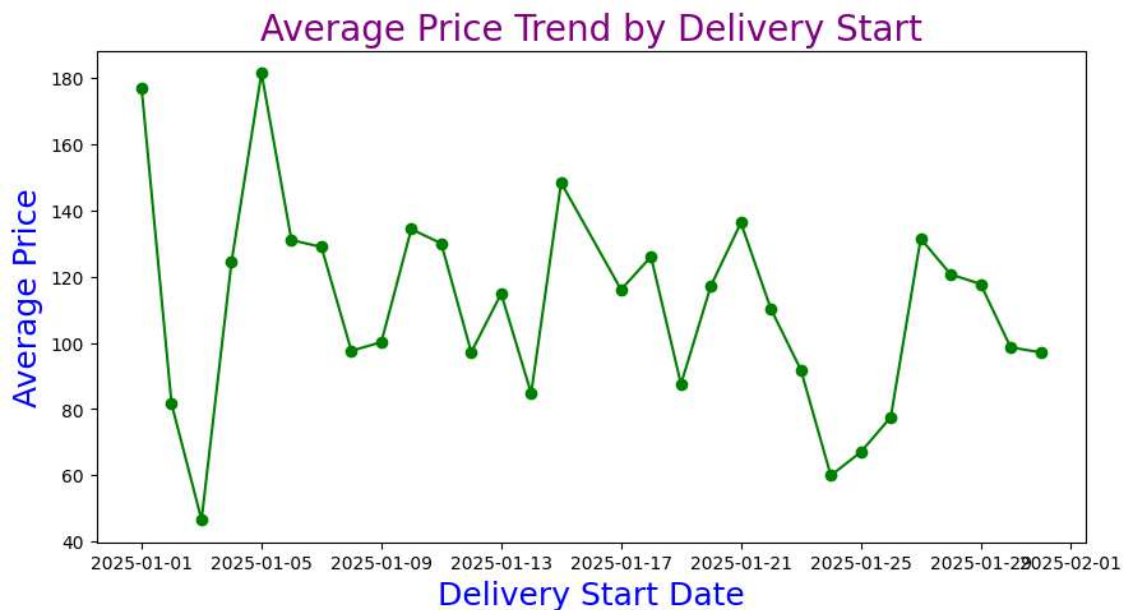
### Explanation

- The pie chart shows that GBP (38%) accounts for the largest portion of trades.
- USD (34%) comes next, indicating its strong role in commodity trading.
- EUR (28%) makes up the smallest share but still represents a substantial portion of trades.

### Takeaway

- The dataset highlights a clear preference for GBP-denominated trades, followed by USD.
- EUR trades, while fewer, remain significant and cannot be ignored in trading or risk analysis.
- For risk managers, this distribution emphasizes the need to monitor GBP and USD exposures closely, as they dominate the trading landscape.

### 5.3 Price Trend by Delivery Start Date (Line Chart)



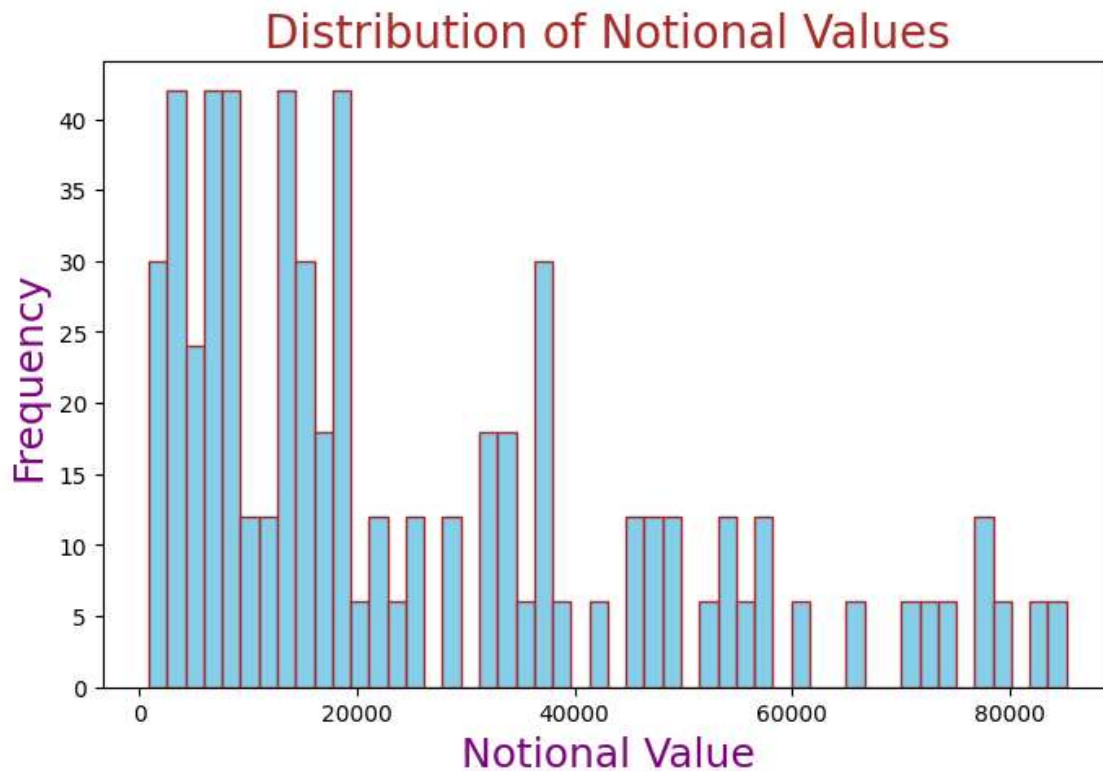
### Explanation

- The line chart shows the average trade price movement during January 2025.
- Prices fluctuated sharply, ranging from as low as ~45 to as high as ~180.
- Several spikes and drops occurred within short intervals, reflecting unstable market behavior.

### Takeaway

- The dataset indicates high volatility in January 2025, with wide price swings in a short period.
- Such fluctuations suggest uncertainty in the commodity market, possibly driven by demand–supply shocks or external factors (e.g., policy, weather, geopolitical events).
- For risk managers, this underscores the importance of proactive hedging strategies (e.g., futures, options) to protect against sudden adverse price movements.

## 5.4 Notional Value Distribution

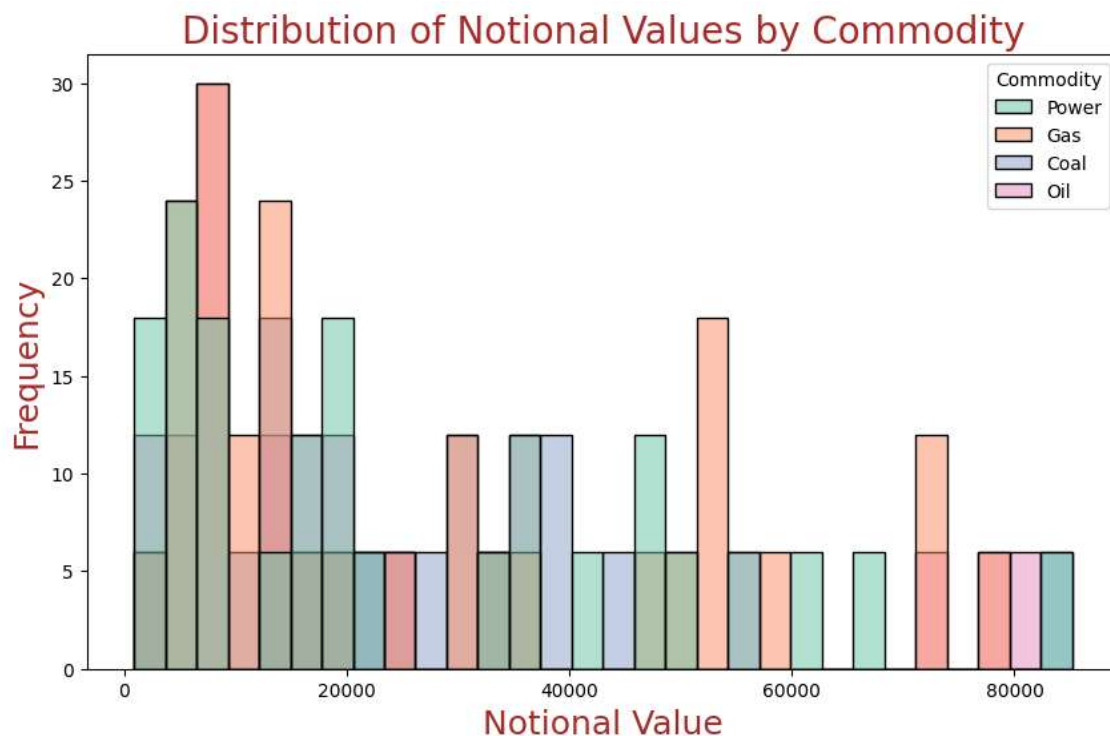


### Explanation

- The histogram shows the frequency of trades across different notional value ranges.
- Most trades fall into the lower notional value bands (below 20,000).
- Higher notional value trades (above 40,000–80,000) exist but are less frequent.
- A few very high-value trades create long tails in the distribution.

### Takeaway

- The market is dominated by smaller trades, which are more frequent and likely represent routine trading activity.
- Large-value trades are rare, but when they occur, they significantly increase market exposure.
- For risk managers, this suggests that while overall market risk is spread across many small trades, occasional high-value trades can create concentrated risk spikes that require careful monitoring.



#### Explanation

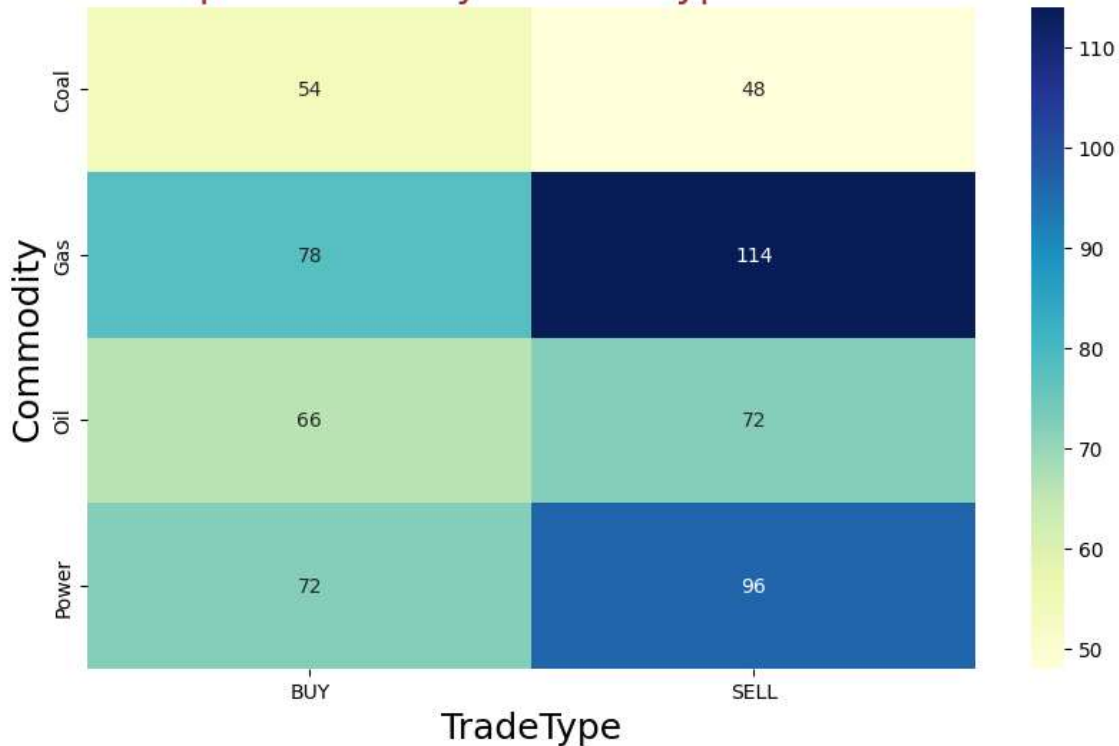
- The histogram breaks down notional values across commodities (Power, Gas, Coal, Oil).
- Most trades in all commodities cluster at lower notional values (<20,000).
- Gas and Power have a relatively wider spread, extending into higher notional value ranges (40,000–70,000).
- Coal shows moderate mid-range notional values (20,000–40,000).
- Oil trades appear less frequent overall, but it still shows presence in both low and high-value ranges.

#### Takeaway

- Small notional value trades dominate across all commodities, confirming that the market is primarily driven by smaller deals.
- Gas and Power exhibit more frequent higher-value trades, suggesting they contribute more to large exposures.
- Coal is more balanced in mid-range trades, while Oil shows fewer but more widely distributed trades.
- For risk managers, Gas and Power require closer monitoring since their notional value spread increases the potential for significant risk spikes.

## 5.5 Commodity vs TradeType (Heatmap)

Heatmap: Commodity vs TradeType Trade Counts



### Explanation

- Gas: 78 BUY trades vs 114 SELL trades → more selling than buying.
- Coal: 54 BUY trades vs 48 SELL trades → slightly more buying.
- Oil: 66 BUY trades vs 72 SELL trades → relatively balanced.
- Power: 72 BUY trades vs 96 SELL trades → more selling than buying.

### Takeaway

- Gas and Power: Traders are reducing exposure with more SELL trades, possibly anticipating weaker demand or oversupply.
- Coal: Shows a mild bullish tilt with slightly more BUY trades.
- Oil: Stays stable, with BUY and SELL trades almost equal, reflecting a balanced outlook.
- Overall, the dataset shows selective bullishness in Coal, stability in Oil, and bearish sentiment in Gas and Power.