Executive Summary – Exploratory Data Analysis (EDA) of PowerCo Data

This analysis explores the **PowerCo dataset**, which combines client-related information and price data to understand customer behavior, energy consumption, and churn patterns. The primary objectives were to assess data quality, identify patterns in customer consumption, detect anomalies, and derive insights that may help in churn prediction and customer segmentation.

1. Data Overview

- Two datasets were analyzed:
  - **Client Data:** Containing customer details such as contract type, sales channel, subscribed power, consumption levels, and churn status.
  - **Price Data:** Containing energy price details across different periods.
- Initial inspection revealed mixed data types, with categorical, numerical, and temporal features.

2. Data Quality Checks

- **Missing values:** Some features (e.g., `channel_sales`) included custom "MISSING" categories, indicating data cleaning interventions.
- **Duplicates:** Checked and removed where necessary.
- **Nulls:** Minimal, but identified and handled.

3. Descriptive Statistics

- Customer consumption and margin variables displayed **highly skewed distributions**, with long right tails indicating a small subset of customers consuming disproportionately high energy.
- Key statistics showed variability in contract types, sales channels, and power subscriptions.

4. Customer Churn Insights

- Around **10% of customers had churned**, aligning with industry expectations.
- Churn distribution varied across sales channels, with the "MISSING" category showing a churn rate of ~7.6%. This suggests that data completeness in sales channel tracking may have predictive value.

5. Consumption Patterns

- **Annual and monthly consumption distributions** were strongly right-skewed, with outliers evident in high-consumption customers.
- Boxplots confirmed the presence of extreme values, which could impact modeling and forecasting accuracy.

6. Forecasting and Margins

- Variables related to consumption forecasts were also positively skewed, requiring transformations for better modeling.
- **Margins and subscribed power** demonstrated outliers and variance across customers, pointing to heterogeneous usage behavior.

7. Key Insights

- Customer churn is influenced by sales channels and potentially by consumption behavior.
- Data distributions suggest the need for transformations and possibly segmentation to capture customer heterogeneity.
- Outliers in consumption and margins require careful handling to avoid distortion in modeling.
- Features such as **contract type, subscribed power, and channel_sales** appear significant for predictive modeling.