# Energy Price Forecasting Project

This project presents a comprehensive energy forecasting solution using publicly available energy market data. It includes data acquisition, preprocessing, time series modeling using ARIMA and Prophet, and the generation of visualizations and structured CSV outputs to simulate and analyze energy pricing scenarios.

---

## Contents



---

## Data Sources

- **EIA API**: Used to pull natural gas spot and futures prices.
    - Endpoint: `https://api.eia.gov/v2/natural-gas/pri/fut/data/`
- **EIA Excel Downloads**: Used to collect historical wholesale electricity data (2020–2023)

---

## Project Structure

Energy_Forecasting/
├── data/

```
│   ├── spot_pricing.csv
│   ├── futures_pricing.csv
│   ├── WholesaleElectricity2020-23.csv
│   ├── spot_prices_arima_forecast.csv
│   ├── spot_prices_prophet_forecast.csv
│   ├── futures_prices_arima_forecast.csv
│   ├── futures_prices_prophet_forecast.csv
│   ├── wholesale_prices_arima_forecast.csv
│   ├── wholesale_prices_prophet_forecast.csv
├── reports/
│   ├── spot_prices_forecast_plot.png
│   ├── futures_prices_forecast_plot.png
│   ├── wholesale_prices_forecast_plot.png
├── src/
│   ├── get_data.py
│   ├── forecasts.py
│   ├── spot_forecasts.py
│   ├── futures_forecasts.py
│   ├── wholesale_forecasts.py
├── config.py
├── README.md
```

---

# Data Overview

## Natural Gas Spot Prices

- File: `spot_pricing.csv`
- Pulled from the EIA API
- Consists of daily spot prices from 2019 to 2023 for the New York region
- Prices reflect the cost of natural gas available for immediate delivery, calculated based on market transactions at major U.S. hubs

## Natural Gas Futures Prices

- File: `futures_pricing.csv`
- Also pulled from the EIA API (combined endpoint)
- Includes daily natural gas futures prices from 2019 to 2023
- Futures prices represent market expectations for future delivery and are derived from financial contracts traded on the NYMEX (New York Mercantile Exchange)

## Wholesale Electricity Prices

- File: `WholesaleElectricity2020-23.csv`
- Compiled from EIA Excel data across major U.S. hubs
- Includes trade-level information such as:
    - Weighted average price: calculated from all trades executed for the day
    - High/low prices: reflect daily volatility across counterparties
    - Daily volume and participant information

---

# Forecasting Methodology

### ARIMA

- Time series model used to capture autocorrelation and trends in price behavior
- Applied to each dataset using:
    - Differencing (d=1)
    - Auto-regression (p=5)
    - No moving average term (q=0)
- Results include:
    - Forecasted price
    - 95% confidence interval

### Prophet

- Facebook's time series forecasting model, handles seasonality and trends well
- 6-month forecast horizon
- Output includes:
    - Point forecast
    - Upper and lower bounds at 95% confidence level

---

# Output Files

### Forecast CSVs (`data/`)

Each of these contains:

- `date`: Date of forecast or historical point
- `forecast`: Predicted price for that date
- `lower_ci`, `upper_ci`: Lower and upper confidence bounds

Forecasted files:

- `spot_prices_arima_forecast.csv`
- `spot_prices_prophet_forecast.csv`
- `futures_prices_arima_forecast.csv`
- `futures_prices_prophet_forecast.csv`
- `wholesale_prices_arima_forecast.csv`
- `wholesale_prices_prophet_forecast.csv`

### Plots (`reports/`)

Each PNG file includes:

- Historical data
- Forecast lines (ARIMA and Prophet)
- Shaded areas for 95% CI

---

# Column Definitions

### Common Fields in Forecast CSVs

| Column | Description |
|---|---|
| date | Date of prediction or observation |
| forecast | Model's predicted price (USD per unit) |
| lower_ci | Lower bound of 95% confidence interval |
| upper_ci | Upper bound of 95% confidence interval |

### Additional Fields in Wholesale CSV

| Column | Description |
|---|---|
| Price hub | Market region or node name |
| Trade date | Original date of transaction |
| Delivery start/end | Period electricity is delivered |
| High/Low price | Daily high/low price in $/MWh |
| Wtd avg price | Weighted average price across trades |

| | |
|---|---|
| Change | Change from previous trade date |
| Daily volume MWh | Total volume traded in megawatt-hours |
| Number of trades | Number of unique trades that day |
| Number of counterparties | Count of unique participants trading |

### Additional Fields in Natural Gas Spot & Futures CSVs

| Column | Description |
|---|---|
| area-name | Name of the market region or pricing hub |
| process | Type of price (e.g., "Spot", "Futures") |
| value | Observed price in USD per MMBtu |
| period | Date the price was recorded (in UTC) |
| units | Unit of measurement (typically USD/MMBtu) |
| date | Converted datetime field used for modeling |

---

## Scenario Simulation

The forecasts can be used in Power BI or Excel to:

- Simulate price spikes or dips
- Apply ±10% shocks to forecasted data
- Compare commodity types (spot vs futures vs wholesale)
- Visualize model confidence intervals

---

## Insights & Observations

- **Early 2021**: Electricity prices spiked due to extreme weather events including Winter Storm Uri (February 2021), especially in ERCOT and PJM regions.
- **Late 2021–2022**: Elevated natural gas prices linked to global LNG demand rebound post-COVID, as well as geopolitical tension.

- **2022**: Russia's invasion of Ukraine (February 2022) led to global energy supply fears, driving both futures and spot price volatility.
- **2023**: Gradual price stabilization as storage volumes improved and demand moderated with milder winters.
- **Wholesale electricity spikes**: Frequently align with seasonal demand peaks (e.g., summers and winters) and transmission congestion.
- **Summer 2022**: High electricity prices in Western U.S. hubs coincided with heatwaves and drought impacting hydro and solar production.
- **Spring 2023**: Slight downturn in futures pricing due to global inventory build-up and mild weather forecasts.

---

# Usage

1. Run the `get_data.py` script to pull and save the latest natural gas pricing data
2. Run `spot_forecasts.py`, `futures_forecasts.py`, and `wholesale_forecasts.py` to:
    - Fit ARIMA and Prophet models
    - Output forecasted CSVs and plots
3. Load outputs into Power BI or Excel for visualization and scenario analysis

---

# Summary

This project demonstrates data collection, time series forecasting, and structured analysis for three energy pricing categories. The output can support informed decision-making for pricing, hedging, and procurement strategies.