**Data-Driven Analysis of Electricity Generation and Consumption (2018–2023)**

Proposal for Energy Insights, Forecasting & Sustainability Planning

**1. Project Overview**

This proposal aims to extract meaningful insights from a large-scale electricity dataset covering hourly records from 2018 to 2023. The dataset records consumption and generation across 20+ energy sources and currency-adjusted electricity prices (TRY, USD, EUR). The focus is on understanding energy consumption patterns, source dependencies, pricing behavior, and strategic opportunities for enhancing efficiency and sustainability. All insights will be derived using descriptive analytics, classical regression models, and clustering techniques.

**2. Business Questions**

|  |  |
| --- | --- |
| Category | Questions |
| Consumption & Demand | What are the hourly, daily, and monthly patterns in electricity consumption? |
| How does electricity demand differ between weekdays and weekends? |
| Are there significant seasonal trends in electricity usage (e.g., summer vs. winter)? |
| Which year had the highest average electricity consumption, and why? |
| Generation & Sources | What percentage of total electricity generation comes from renewable vs. fossil sources? |
| How has the reliance on each energy source changed from 2018 to 2023? |
| Which renewable source (solar, wind, hydro) has shown the most growth? |
| What is the variability of solar and wind generation throughout the day and year? |
| Demand-Supply Balance | How often does electricity generation exceed or fall short of consumption? |
| During which time periods (hour, month, season) do imbalances most frequently occur? |
| How dependent is the system on imported electricity during peak hours? |
| Pricing & Economics | How do electricity prices correlate with demand levels? |
| Which energy sources are most associated with high pricing periods? |
| Is there a seasonal or annual trend in electricity prices? |
| How do currency fluctuations affect local pricing in TRY? |
| Variable Relationships | What are the strongest predictors of electricity price per MWh? |
| Are there any surprising negative or positive correlations between variables? |
| How do combinations of energy sources impact overall system efficiency and price stability? |
| Clustering & Profiles | **Can we classify days into distinct demand profiles (e.g., working day vs. holiday)?** |
| **What are the common patterns among high-demand or low-demand days?** |
| **Which clusters show the highest reliance on renewable energy?** |
| Policy & Strategy | What scenarios suggest a need for more renewable investment? |
| Where are the most critical inefficiencies or vulnerabilities in the energy mix? |
| What would be the impact of replacing a percentage of fossil generation with renewables? |

**3. Dataset Description**

**Structure:**

* **Timeframe:** January 2018 – December 2023 (hourly records).
* **Total Records:** 52,584
* **Total Columns:** 22
  + **Time:** Timestamp of each record.
  + **Consumption & Generation:**
    - consumption\_MWh.
    - total\_generation\_MWh.
  + **Energy Sources:**
    - Fossil-based: natural\_gas, coal\_imported, fuel\_oil, asphaltite\_coal, hard\_coal, lignite, LNG, naphtha.
    - Renewable: hydro\_dam, hydro\_river, solar, wind, biomass, geothermal, waste\_heat.
    - Others: International (Imports).
  + **Electricity Prices:**
    - TRY/MWh, USD/MWh, EUR/MWh.

**Observations:**

* Clean, continuous hourly coverage (no gaps or nulls).
* Data is well-suited for forecasting, regression analysis, and comparative visualizations.

**4. Objectives**

**A. Trend Analysis**

* Explore electricity demand and generation trends by hour, day, and month.
* Analyze seasonal and weekly patterns (e.g., weekends vs. weekdays).
* Investigate long-term changes in usage behavior from 2018 to 2023.

**B. Source Dependency Analysis**

* Determine each energy source’s average share in electricity generation.
* Identify shifts in fuel mix (e.g., increasing renewables, declining coal).
* Evaluate the stability and intermittency of renewable sources (e.g., solar, wind).

**C. Demand-Supply Balance**

* Compare consumption\_MWh with total\_generation\_MWh to detect:
  + Overproduction.
  + Deficits requiring import support.
* Assess the impact of source composition on grid stability.

**D. Price Dynamics**

* Correlate electricity prices in TRY, USD, and EUR with:
  + Total demand.
  + Individual source contributions.
  + Use of imported energy.
* Identify major drivers of high/low pricing in TRY, USD, and EUR.

**E. Data-Driven Modeling**

* Build regression models to:
  + Predict electricity prices (TRY/MWh) based on energy mix and demand.
  + Identify most impactful features (natural gas, imports, solar, etc.).
* Apply clustering to:
* Detect typical usage profiles (e.g., peak-load days vs. low-demand days).
* Group days by energy composition (fossil-heavy vs. renewable-heavy).

**5. Proposed Methodology**

**Data Preparation**

* Clean and parse datetime values.
* Engineer new features:
* hour, day\_of\_week, month, season.
* % share of each generation source.
* Handle outliers and scaling for modeling.

**Exploratory Data Analysis (EDA)**

* **Heatmaps** of average usage by hour/week/month.
* **Stacked area plots** showing source contributions over time.
* **Scatterplots and pairplots** for price correlation.
* **Boxplots** for seasonal variations in pricing and generation.

**Modeling Techniques**

* **Regression Models:**
  + Linear Regression.
  + Lasso/Ridge (for feature selection and multicollinearity control).
  + Random Forest Regression.
* **Feature Importance Analysis:**
  + Assess which energy sources most influence pricing and balance.
* **Clustering Analysis:**
  + Use K-Means or Hierarchical Clustering to discover energy usage profiles.

**6. Deliverables**

|  |  |  |
| --- | --- | --- |
| Component | Description | |
| 📊 Interactive Dashboard | Statistical insights and charts | |
| 📈 EDA Reports | Python scripts + trained models for demand & price predictions | |
| 🧠 Forecasting Models | Regression models for price analysis | |
| 📄 Strategic Summary | | Clear recommendations and findings for stakeholders |

**7. Potential Impacts**

* Reveal inefficiencies in the current energy mix.
* Empower decision-makers with actionable insights for reducing costs.
* Highlight key leverage points for investing in renewable sources.
* Inform national energy planning and pricing regulation strategies.

**8. Tools and Technologies**

|  |  |
| --- | --- |
| Tool | Purpose |
| Python (Pandas, NumPy) | Data handling and transformation |
| Matplotlib, Seaborn, Plotly | Visualization |
| Scikit-learn | |  | | --- | | Regression modeling and clustering |  |  | | --- | |  | |
| Streamlit | Reporting and dashboards |
| Colab Notebook | |  | | --- | | Workflow and documentation |  |  | | --- | |  | |

**8. Conclusion**

This project outlines a clear and practical approach to analyzing electricity generation, consumption, and pricing data between 2018 and 2023. By leveraging a rich, hourly dataset and applying descriptive analysis, regression modeling, and clustering techniques, the study aims to deliver meaningful insights into how energy is produced, consumed, and priced over time.

Key questions around demand patterns, source dependency, supply balance, and price dynamics will be addressed using classical statistical tools. The analysis will not only help identify inefficiencies in the current energy mix but also highlight which sources drive pricing volatility and when the system relies heavily on imports or fossil fuels.

The results of this project will serve as a valuable resource for energy planners, policymakers, and sustainability stakeholders, enabling data-driven decisions that enhance reliability, reduce costs, and support the shift toward a cleaner and more balanced energy future.

**9. Data Source**

<https://www.kaggle.com/code/pythonafroz/eda-on-energy-deficit/notebook>