Analysis and Optimization of Energy Consumption

Instructions

- The **project** must be sent no later than the **25th of March, 10am**. This includes :
  - The **specifications document**.
  - The project as in the **code and/or dashboard**, with his **sources**.
    - Any project not sent with the associated sources will be considered not sent.
  - The **presentation for your deliverables** can be sent directly after your oral presentation, but no later than 1 hour beyond that.
  - Past the deadline, the overall grade of your project will be penalized.

- An **intermediary deliverable** must be sent by **06th of March, 10am**.
  - This deliverable should act as a **waypoint** in your work and allow the jury to appreciate the way you integrated their feedback into the final material.
  - This deliverable will not be graded but its absence will **impact your organization grade and thus your overall grade.**

- Each deliverable needs to be sent to :
  ambre.lepeudry@mc2i.fr, anthony.foulon@mc2i.fr, quentin.blaclard@mc2i.fr, amelie.gerard@mc2i.fr, sybille.dessart@mc2i.fr, marc.gourvitch@mc2i.fr, nicolas.legrand@mc2i.fr, edgar.sorlin@mc2i.fr, nedra.mellouli@devinci.fr

  Your oral presentation will be **45 minutes long per group**.
  This should leave you with **30 minutes of présentation** and **15 minutes for questions/answers.**
  You group must be composed of at least 4 persons.

Notation

Specifications document & Project			Oral presentation (group grade)		Oral presentation (individual grade)	
Organization	Pertinence	Legibility	Group cohesion	Coherence of speech	Answers	Speech clarity
20%	60%	20%	40%	60%	75%	25%
50%			25%		25%	

Context

Against a backdrop of energy transition and global warming, companies and organizations need to reduce their environmental impact while optimizing their resources. Controlling energy consumption is a strategic factor in meeting these global challenges.

The aim of this project is to give you the opportunity to work on real-life problems linked to energy analysis, mobilizing your skills in data management, predictive modeling and operational recommendations.

Challenges and needs :

1. **Analyze energy consumption :**
    - Study the existing situation to identify trends and consumption periods.
    - Identify and analyze peak consumption periods.
    - Understand the causes of energy peaks through correlations with other data (e.g. meteorological data).
2. **Anticipate and model :**
    - Develop models to forecast energy consumption.
    - Quantify potential savings under different scenarios.
    - Anticipate periods of tension on the energy network.
3. **Optimize operations :**
    - Identify optimization levers based on analyzed data.
    - Propose concrete actions, prioritized according to potential gains.
    - Measure and assess the impact of actions implemented.

Technical requirements :

Data analysis :

- Knowledge of data manipulation tools and languages, such as Python (Pandas, NumPy ...)
- Basics of data cleaning and preparation (handling missing values, managing outliers, etc.).
- Ability to explore data to identify trends and anomalies.

Data visualization :

- Competence in the use of visualization libraries such as Matplotlib, Seaborn, Plotly, or tools such as Power BI or Tableau.
- Create clear, informative graphics adapted to different audiences (technical or non-technical).

Predictive modeling and machine learning :

- Knowledge of the basics of machine learning (supervised and unsupervised algorithms).
- Use libraries like Scikit-learn to build predictive models.
- Understanding of the concepts of cross-validation and model evaluation (accuracy, RMSE, etc.).

Expected work

Project phases :

➔ Phase 1: Data collection and preparation

Objective: Prepare a clean, usable dataset for analysis and modeling.

1. **Data collection** :
    - Identify and gather data from a variety of accessible sources (e.g. internal company data, public data, IoT sensors, weather data ....).
    - The choice of data sources is free, but they must be relevant to the analysis of energy consumption.
2. **Cleaning and Data Preparation** :
    - Manage missing values, inconsistencies and anomalies (e.g. outliers).
    - Standardize formats (dates, consumption units, etc.) to guarantee data consistency.
    - Create new variables if necessary, such as consumption periods (off-peak, peak, seasonal, etc.).

➔ Phase 2: Analysis and Modeling

Objective: Explore the data and create models adapted to the issues identified.

1. **Exploratory data analysis (EDA**) :
    - Identify main trends, consumption patterns, anomalies and correlations with other variables (e.g. temperature, days of the week, special events).
    - Visualize these results in simple, relevant graphs (trend curves, histograms, scatter diagrams, etc.).
2. **Modeling** :

- **Predictive model**: Build a model that predicts energy consumption based on the factors identified.
- **Anomaly detection**: Develop a model to identify atypical consumption periods or energy peaks.
- **Classification**: Identify consumption periods (e.g. days with high consumption, peak hours).
- Students can choose from several techniques: regression, clustering, time series, or others depending on the nature of the data.

3. **Model Evaluation and Optimization** :
   - Test and compare model performance using appropriate evaluation criteria (e.g. accuracy, RMSE, etc.).
   - Optimize models to improve accuracy or relevance of results.

➔ **Phase 3: Visualization and presentation of results**

Objective: Communicate results clearly and concisely.

1. **Creating a dashboard** :
   - Develop an interactive dashboard to visualize the results of analyses and models, for example :
     - Estimated energy consumption.
     - Detection of consumption peaks.
     - Classification or clustering results.
   - Free choice of visualization tools (e.g. Tableau, Power BI, etc.).
2. **Preparing the final presentation** :
   - **Context and challenges**: Explain business needs and energy consumption issues.
   - **Methodological approach**: Describe the steps taken, choice of tools, data collection, preparation, analyses performed and models developed...
   - **Presentation of results and insights**
   - **Recommendations** : Propose concrete actions based on the results of the analyses, as optimization levers for managing energy consumption.

Deliverables Expected :

1. **Prepared dataset**: data cleaned, structured and ready for analysis.
2. **Prediction models and analysis**: Code or notebooks with developed models and their evaluation.
3. **Dashboard / Visualizations**: Interactive graphs and dashboards to visualize key results.
4. **PowerPoint presentation**: Project summary, methodology, results and recommendations.

Means & Tools

- **Public data sources**
  - Open Data Réseaux Énergies (ODRÉ): https://opendata.reseaux-energies.fr/
  - Base SDES - Ministère de la Transition écologique : Home | Statistical data and studies on climate change, energy, environment, housing and transport
  - Meteorological Data Sources

    - **Météo** France: https://donneespubliques.meteofrance.fr/
    - OpenWeatherMap: https://openweathermap.org/api
    - ERA5: https://cds.climate.copernicus.eu/
    - Kaggle Energy Datasets: https://www.kaggle.com/datasets?search=energy+consumption
- **Development tools :**
    - **Python and libraries**:
        - Pandas: https://pandas.pydata.org/
        - NumPy: https://numpy.org/
        - Scikit-learn: https://scikit-learn.org/
        - XGBoost: https://xgboost.readthedocs.io/
        - LightGBM: https://lightgbm.readthedocs.io/
        -
- **Visualization tools :**
    - **Plotly**
        - Main site: https://plotly.com/python/
        - Documentation: https://plotly.com/python/getting-started/
    - **Power BI**
        - Main site: https://powerbi.microsoft.com/
        - Tutorials: https://learn.microsoft.com/fr-fr/power-bi/

- **Development Environment**
    - **Jupyter Notebook**
        - Main site: https://jupyter.org/
        - Anaconda: https://www.anaconda.com/download
        - Google Colab: https://colab.research.google.com/
    - **VS Code**
        - Download: https://code.visualstudio.com/
        - Python Extension: Extension Marketplace
        - Jupyter Extension: Extension Marketplace
    - **Git and GitHub**
        - Git: https://git-scm.com/downloads
        - GitHub Desktop: https://desktop.github.com/
        - Beginner's guide: https://github.com/git-guides

Evaluated skills

Data Collection & Integration : Ability to gather, clean and integrate datasets from multiple sources in a single tool. Some data modelization would be appreciated but is optional.
Data Visualization : Capacity to create and design a clear, actionable and visually engaging dashboard with tools such as Power BI, Tableau or others.
Data Analysis : Ability to understand and analyse trends, patterns and their intersection between the needs of the business.
Machine Learning & Predictive Modeling : Ability to apply machine learning techniques and build predictive models to uncover insights, improve decision-making, or automate processes.

Business Communication : Capacity to communicate data findings to a business audience in a clear and concise manner.