ENERGY PROJECT

DATASCIENTIST COURSE 2024

MARYAM MORADI
SARA CERRETO
LEENA WARUNKAR

OVERVIEW OF THE ENERGY PROJECT & DATA FOUNDATIONS

Project Overview:

- **Initiative Focus:** Strategic analysis and optimization of energy consumption in the context of technological advancements and sustainable development.
- **Main Objective:** Harmonize national and departmental energy consumption and production insights, mitigate blackout risks, and forecast consumption with precision.
- **Analytical Approach:** Comprehensive validation process including exploration, visualization, modeling, and transparent reporting.

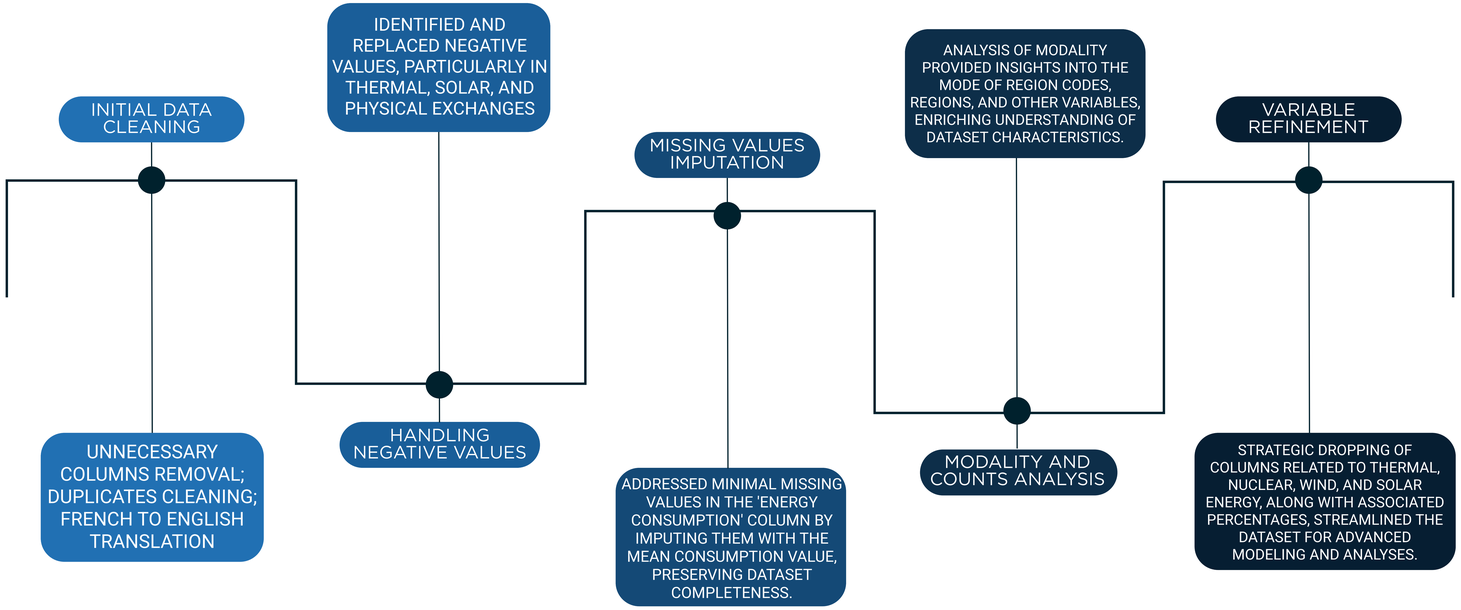
Data Source:

- **Primary Data:** Open Data Energy Networks (ODRE) dataset providing granular, half-hourly insights on energy consumption and production since January 2013.
- **eCO2mix Regional Dataset:** Daily updated data offering a verified, comprehensive view of regional energy dynamics (consumption, diverse production metrics, STEP operational details).
- **Temperature Data:** Regional daily temperature data (since January 2016) to analyze meteorological impacts on energy consumption.
- **Peak Consumption Data:** Historical records of maximum power required for peak electricity consumption, highlighting influences of weather and calendar on energy demand.

OPTIMIZING DATA INTEGRITY: PREPROCESSING INSIGHTS AND CLEANUP ANALYSIS

DATASET BEFORE PREPROCESSING

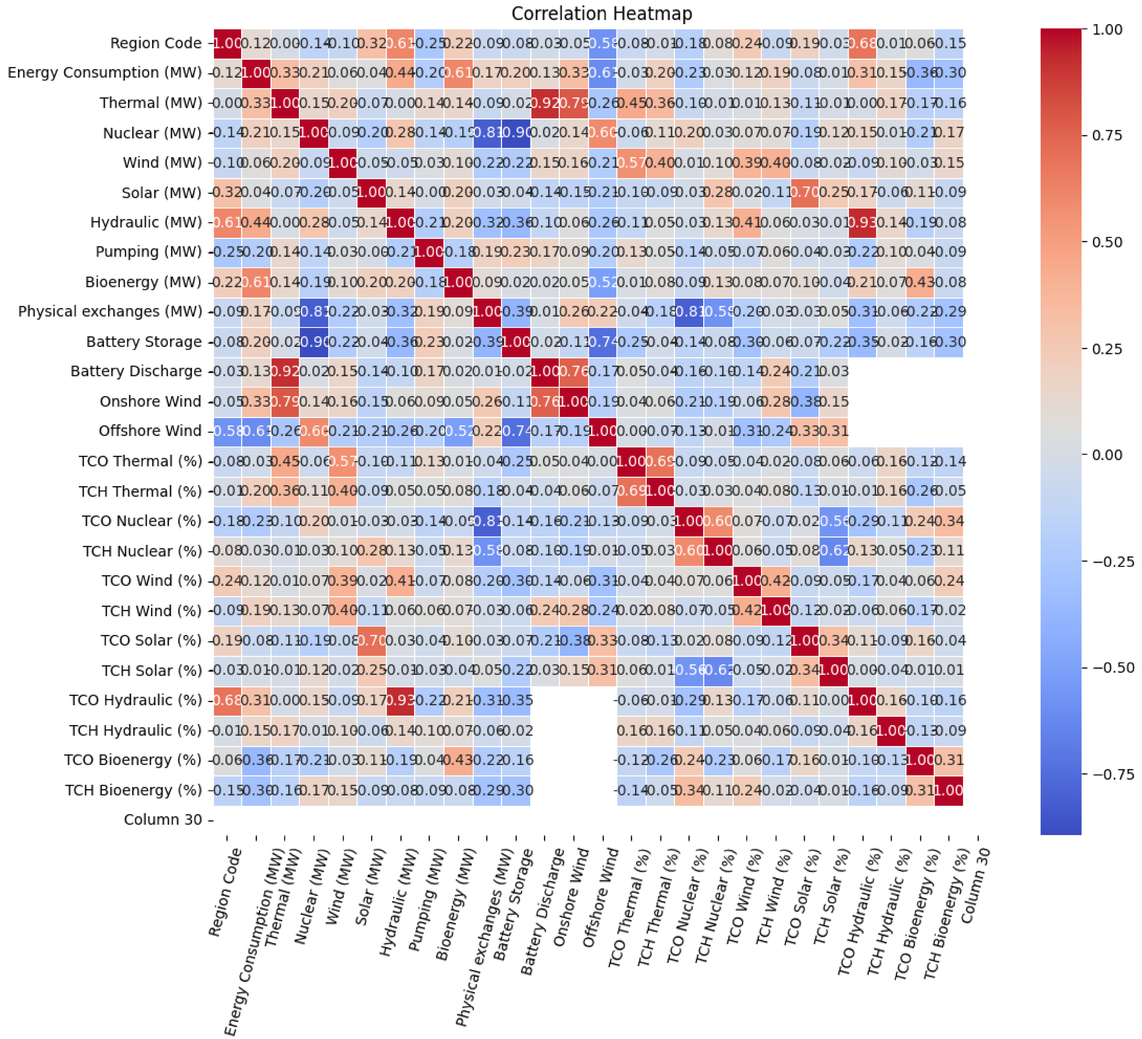
Name of the Column	Variable type	Percentage of missing values	Categorical / Quantitative
Code INSEE région	int64	0.000000	Quantitative
Région	object	0.000000	Categorical - more than 10 categories
Nature	object	0.000000	Unique Value
Date	object	0.000000	Categorical - more than 10 categories
Heure	object	0.000000	Categorical - more than 10 categories
Date - Heure	object	0.000000	Categorical - more than 10 categories
Consommation (MW)	float64	0.008542	Quantitative
Thermique (MW)	float64	0.009254	Quantitative
Nucléaire (MW)	float64	41.671709	Quantitative
Eolien (MW)	float64	0.043423	Quantitative
Solaire (MW)	float64	0.009254	Quantitative
Hydraulique (MW)	float64	0.009254	Quantitative
Pompage (MW)	float64	50.004627	Quantitative
Bioénergies (MW)	float64	0.009254	Quantitative
Ech. physiques (MW)	float64	0.009254	Quantitative
Stockage batterie	float64	100.000000	Quantitative
Déstockage batterie	float64	100.000000	Quantitative
Eolien terrestre	float64	100.000000	Quantitative
Eolien offshore	float64	100.000000	Quantitative
TCO Thermique (%)	float64	100.000000	Quantitative
TCH Thermique (%)	float64	100.000000	Quantitative
TCO Nucléaire (%)	float64	100.000000	Quantitative
TCH Nucléaire (%)	float64	100.000000	Quantitative
TCO Eolien (%)	float64	100.000000	Quantitative
TCH Eolien (%)	float64	100.000000	Quantitative
TCO Solaire (%)	float64	100.000000	Quantitative
TCH Solaire (%)	float64	100.000000	Quantitative
TCO Hydraulique (%)	float64	100.000000	Quantitative
TCH Hydraulique (%)	float64	100.000000	Quantitative
TCO Bioénergies (%)	float64	100.000000	Quantitative
TCH Bioénergies (%)	float64	100.000000	Quantitative

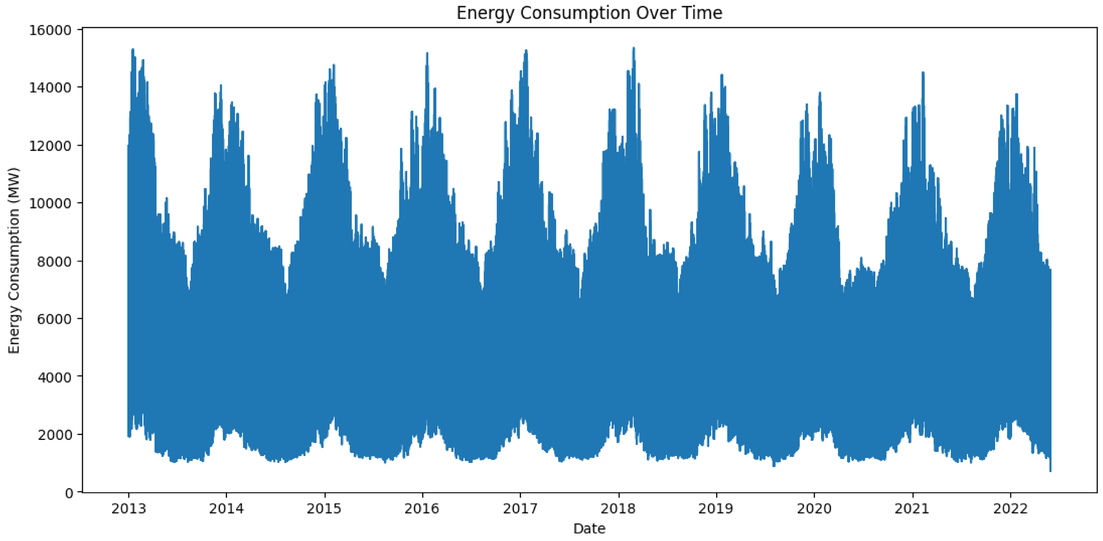


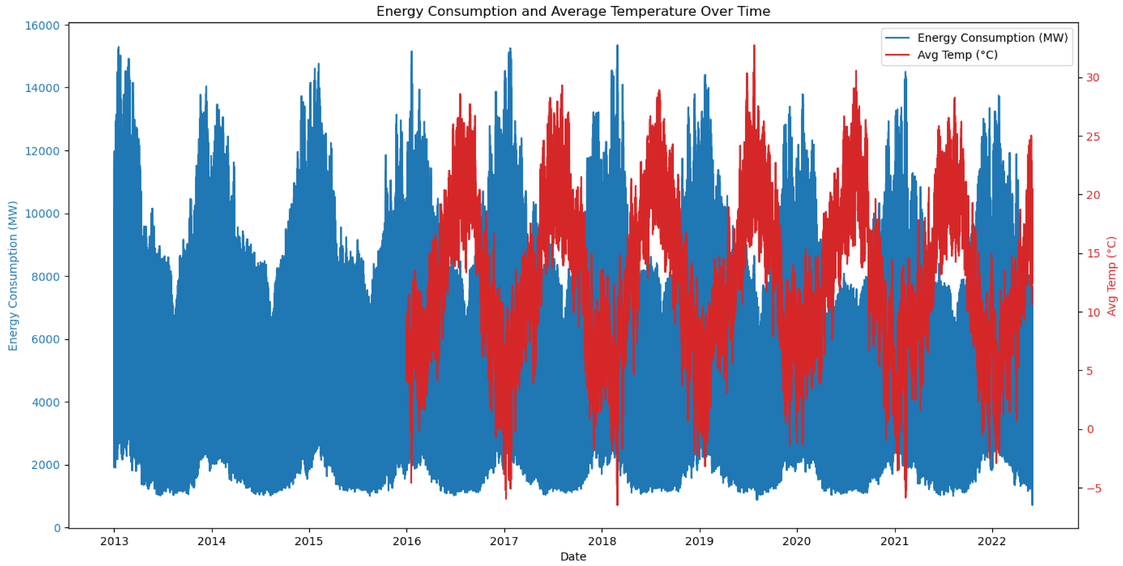
DATASET AFTER PREPROCESSING

# Column	Name of the column in English	Variable type	Percentage of missing values	Categorical / Quantitative
1	Region Code	int64	0.000000	Quantitative
2	Region	object	0.000000	Categorical - more than 10 categories
3	Nature	object	0.000000	Unique Value
4	Date	datetime64[ns]	0.000000	Categorical - more than 10 categories
5	Hour	datetime64[ns]	0.000000	Categorical - more than 10 categories
6	Date - Hour	datetime64[ns]	0.000000	Categorical - more than 10 categories
7	Energy Consumption (MW)	float64	0.000000	Quantitative

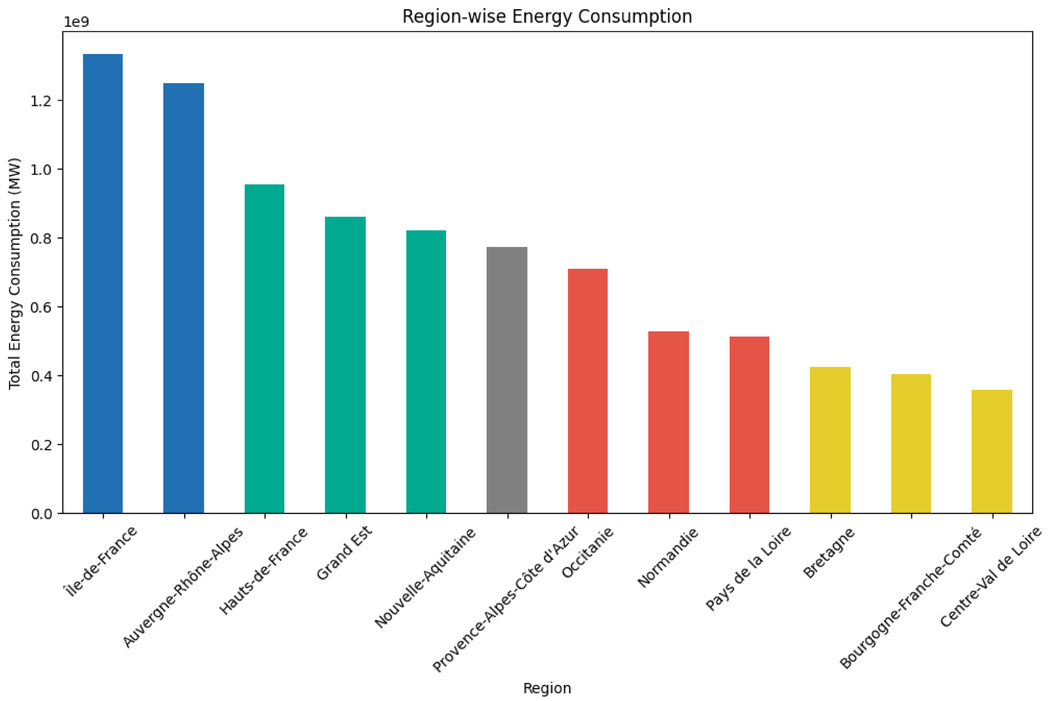
ENERGY CONSUMPTION AND AVERAGE TEMPERATURE OVER TIME

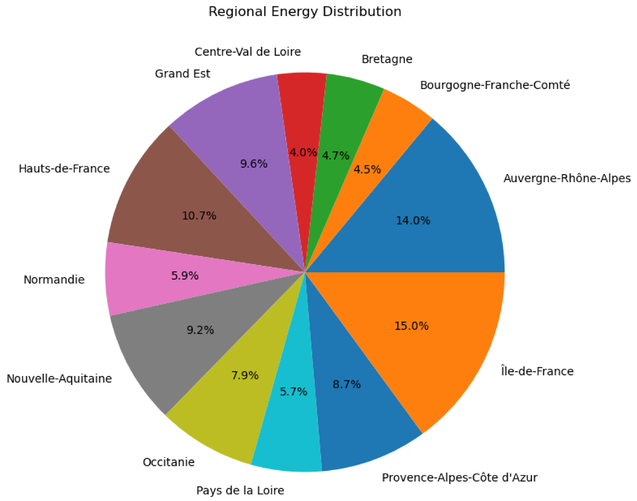


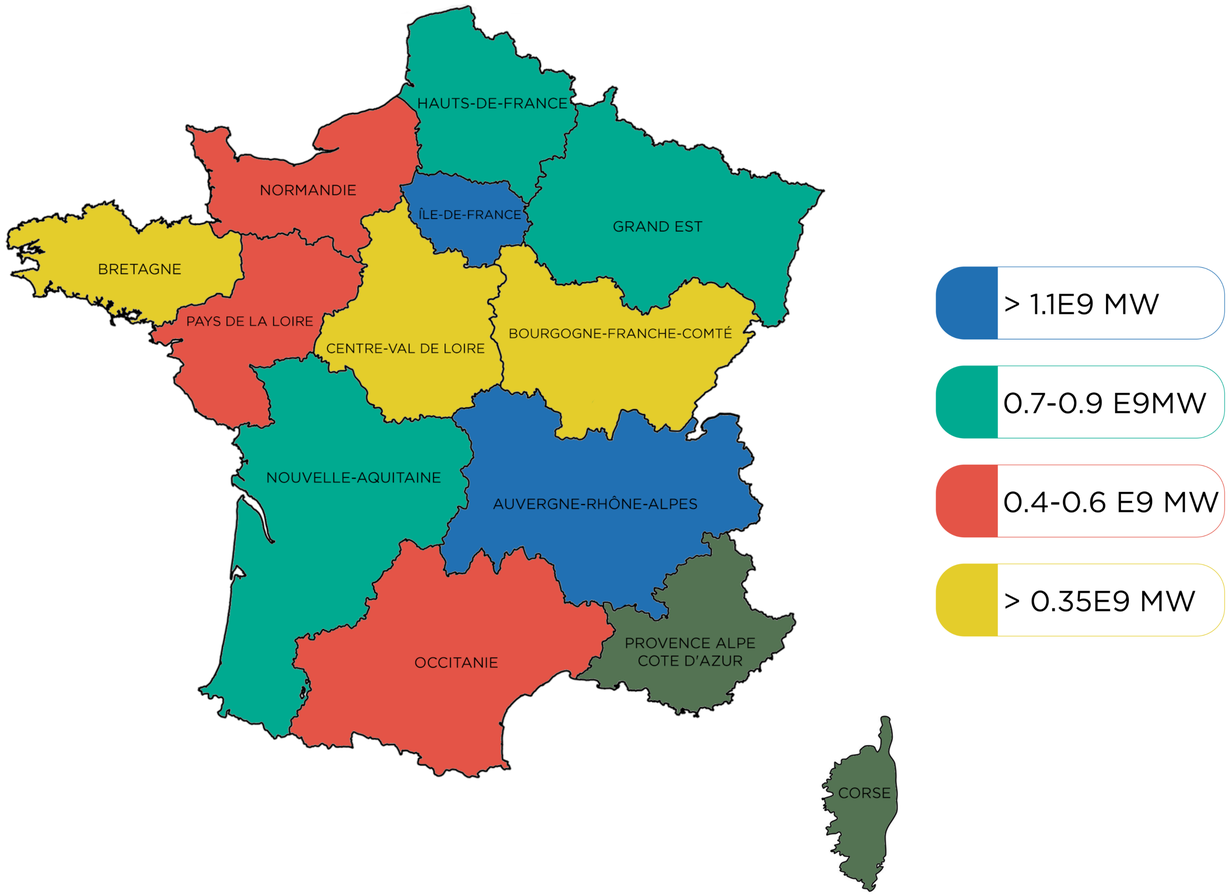




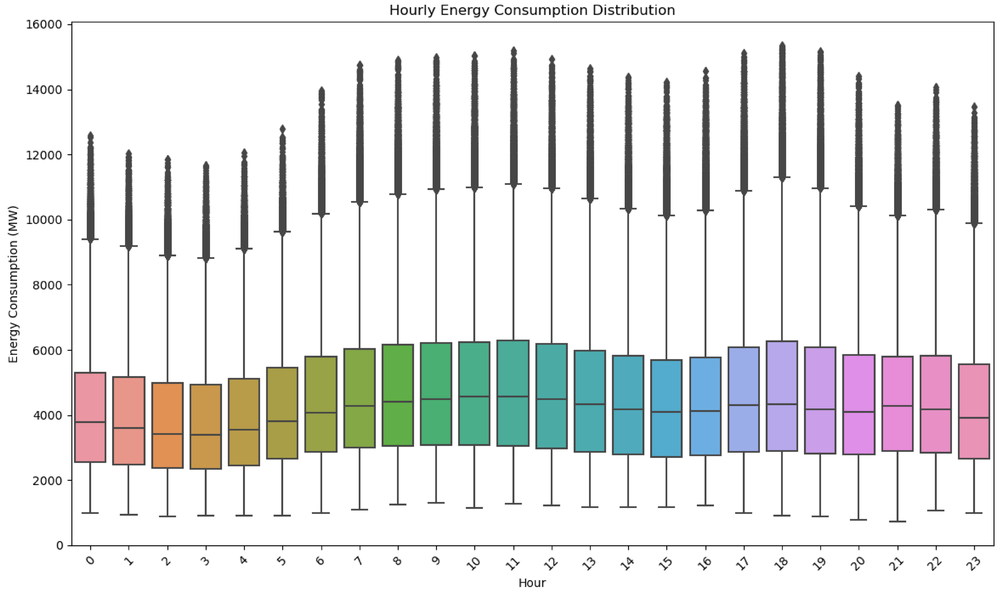
REGION-WISE ENERGY CONSUMPTION

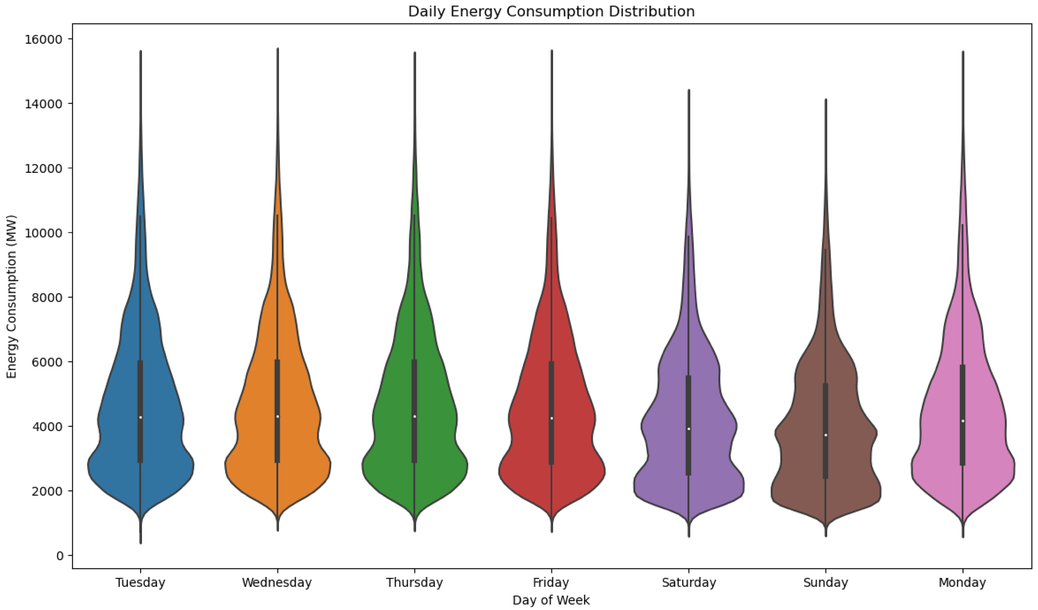


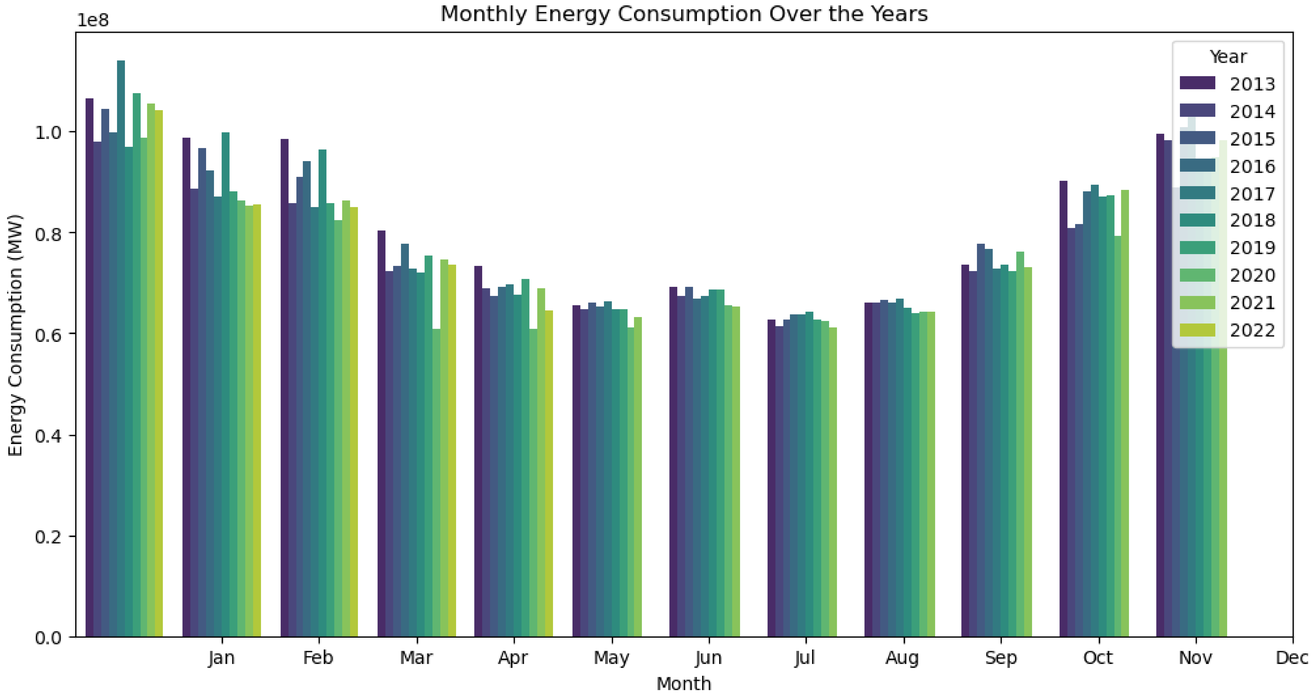




ENERGY CONSUMPTION TRENDS: HOURLY, DAILY, MONTHLY



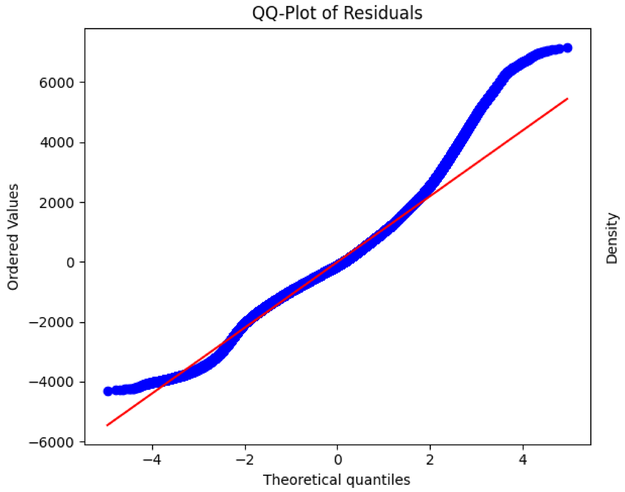


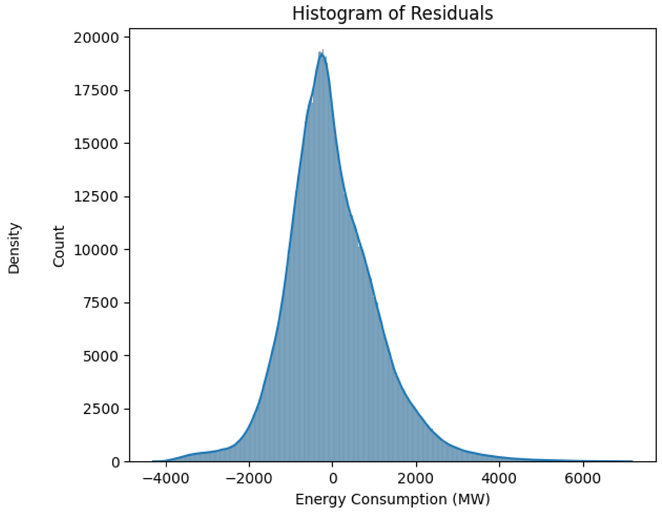


ANALYSIS OF RESIDUALS

ANALYSIS OF VARIANCE (ANOVA) TEST

	df	sum_sq	mean_sq	F	PR(>F)
Region	11.0	6.87	6.24	503734.80	0.0
Residual	1980276.0	2.45	1.24	NaN	NaN

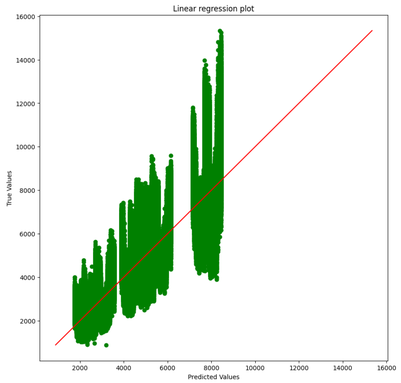


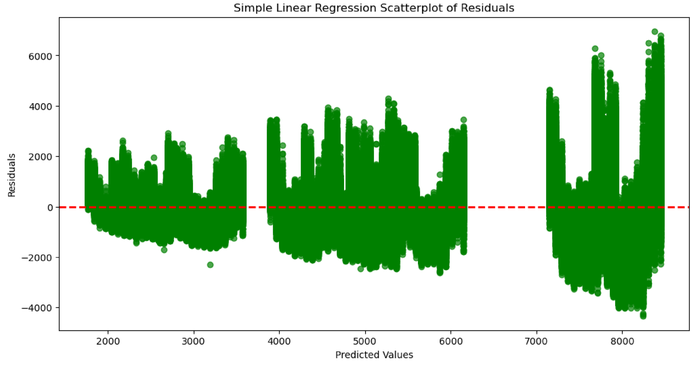


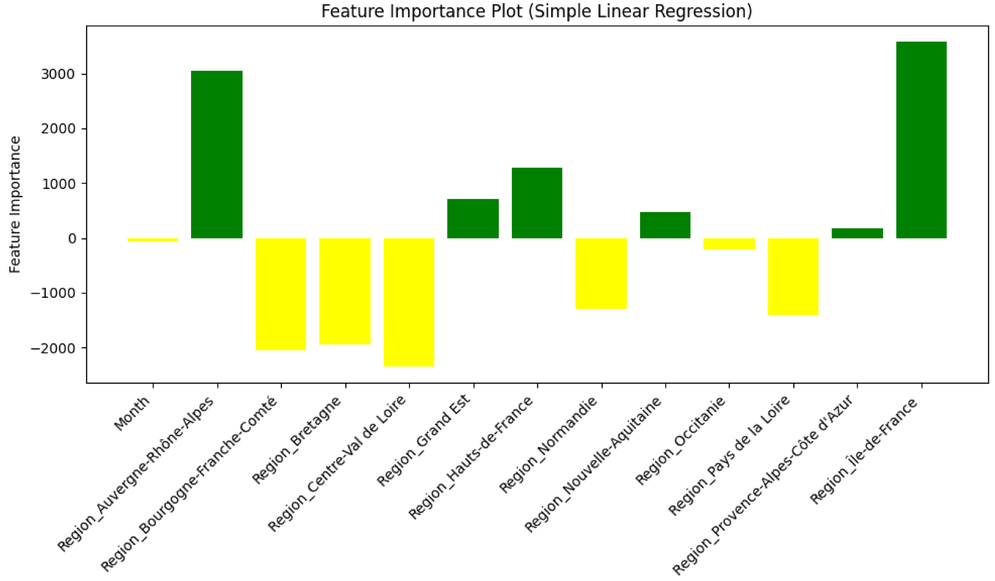
SHAPIRO-WILK TEST AND P-VALUES

#	Region	Shapiro Statistic	P-Value
0	Centre-Val de Loire	0.96	0
1	Île-de-France	0.97	0
2	Grand Est	0.98	0
3	Pays de la Loire	0.97	0
4	Provence-Alpes-Côte d'Azur	0.97	0
5	Normandie	0.97	0
6	Occitanie	0.96	0
7	Bourgogne-Franche-Comté	0.98	0
8	Hauts-de-France	0.98	0
9	Nouvelle-Aquitaine	0.96	0
10	Auvergne-Rhône-Alpes	0.97	0
11	Bretagne	0.97	0

REGRESSION MODELS OVERVIEW: SIMPLE AND MULTIPLE LINEAR

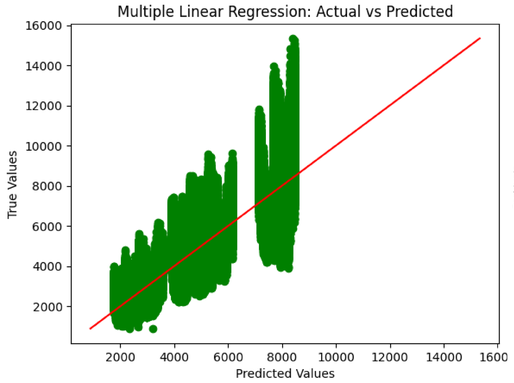


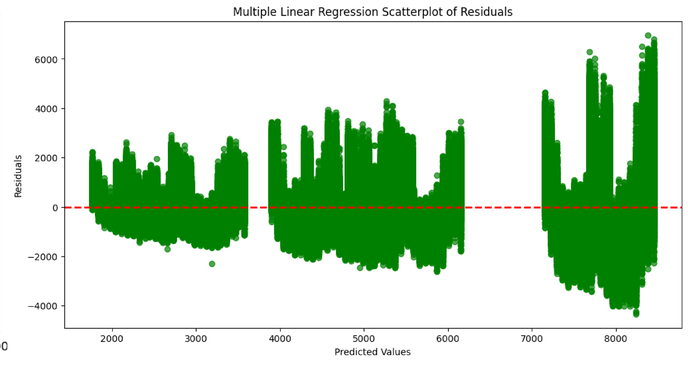


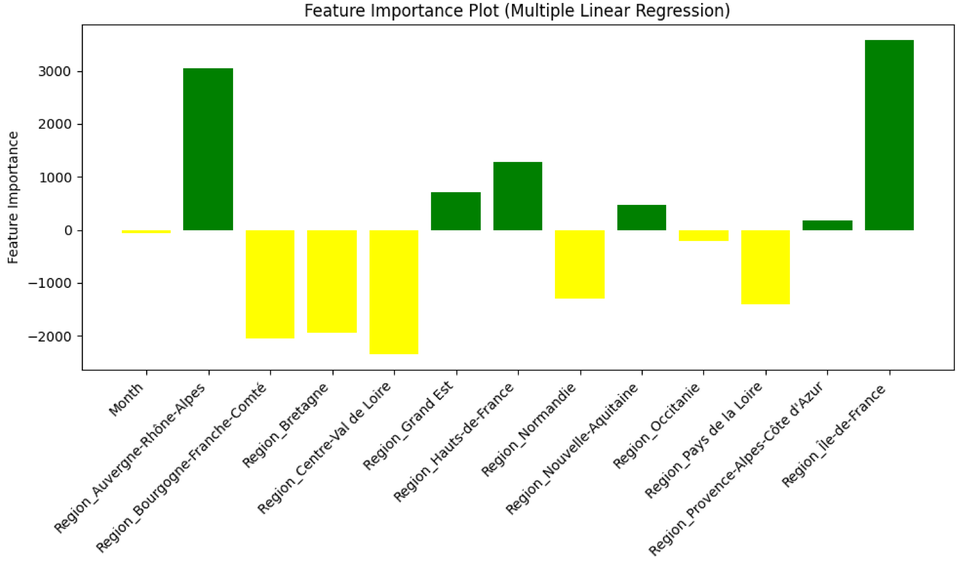


SIMPLE LINEAR REGRESSION MODEL

MEAN SQUARED ERROR: 1182893.2478333903
R² (TRAIN): 0.748985455589243
R² (TEST): 0.7495365698831171



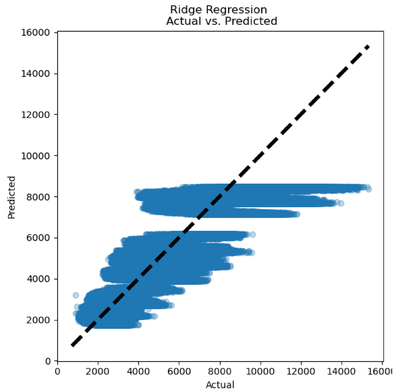


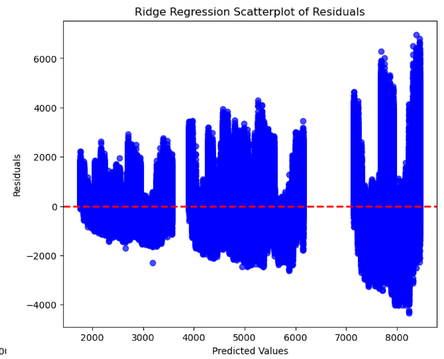


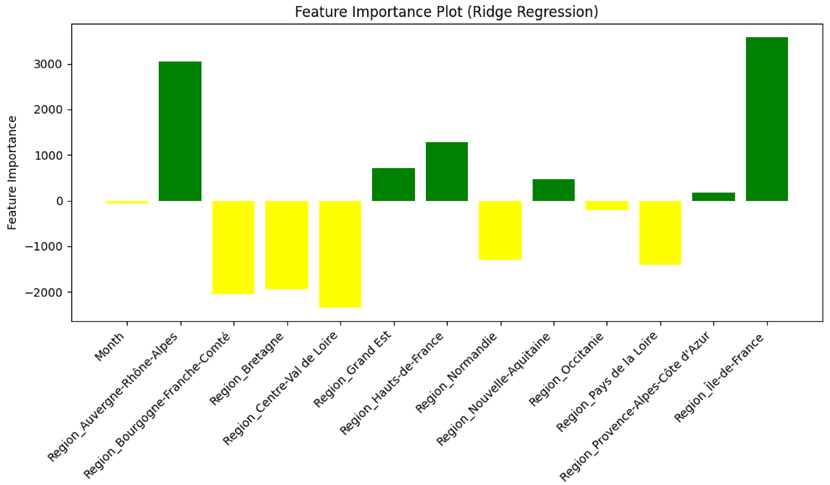
MULTIPLE LINEAR REGRESSION MODEL

MEAN SQUARED ERROR: 1182893.2478333903
R² (TRAIN): 0.748985455589243
R² (TEST): 0.7495365698831171

REGRESSION MODELS OVERVIEW: RIDGE, LASSO, AND RANDOM FOREST

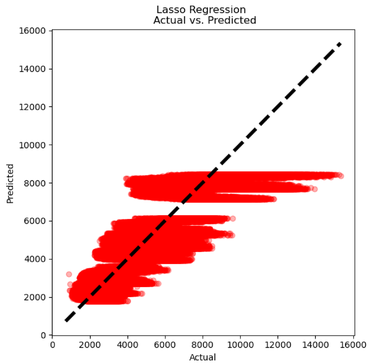


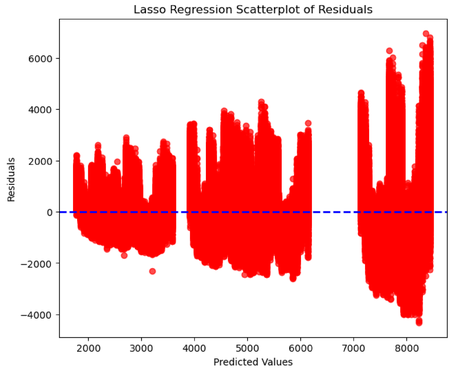


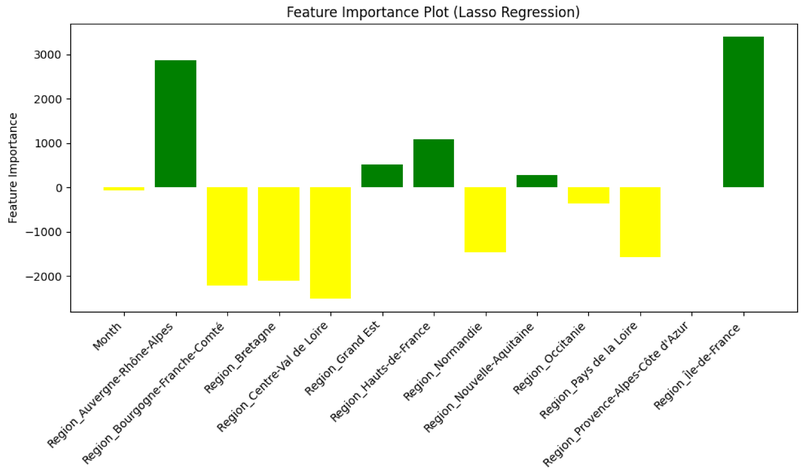


RIDGE REGRESSION MODEL

MEAN SQUARED ERROR: 1182893.323
R^2 (TRAIN): 0.7489854555
R^2 (TEST): 0.7494913754

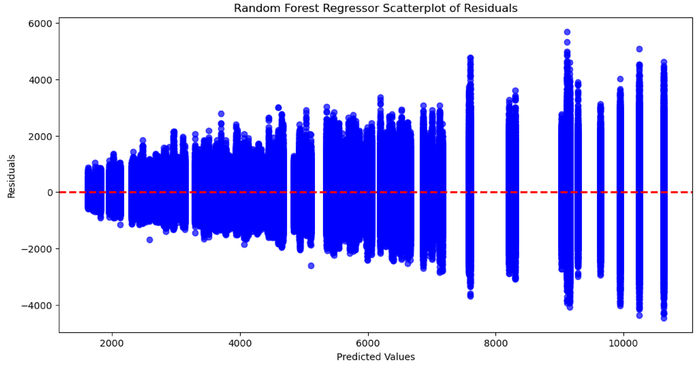


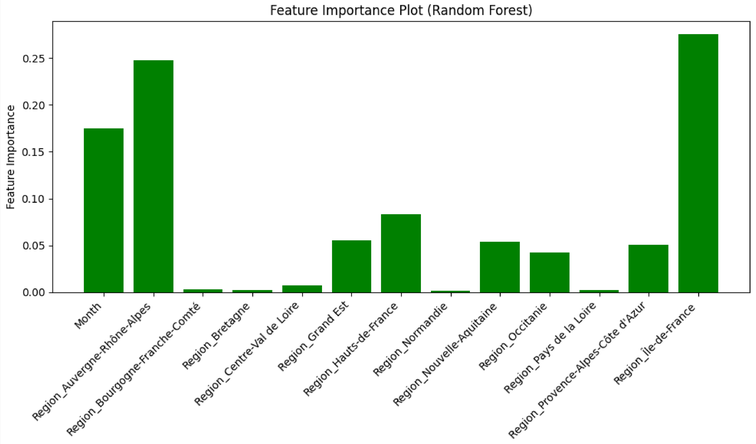




LASSO REGRESSION MODEL

MEAN SQUARED ERROR: 1183106.693
R^2 (TRAIN): 0.7489552977
R^2 (TEST): 0.7494913754





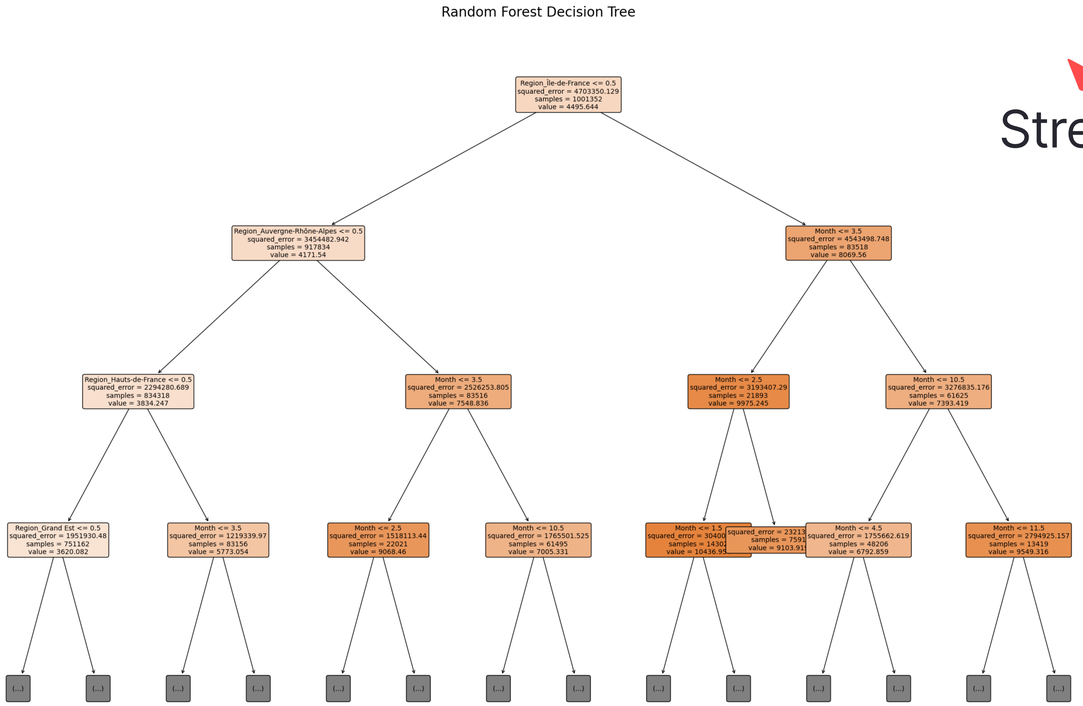
RANDOM FOREST REGRESSION MODEL

MEAN SQUARED ERROR: 497902.83
R-SQUARED (TRAIN): 0.89
R-SQUARED (TEST): 0.89

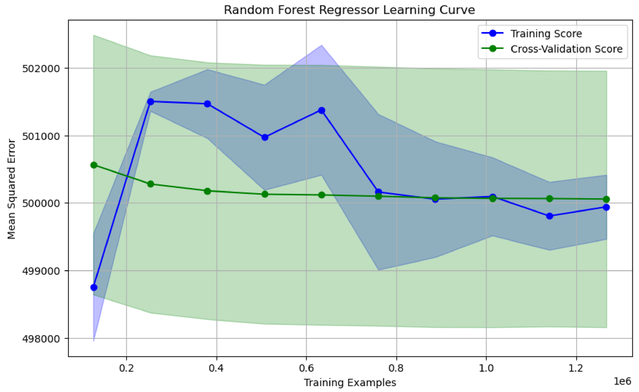
COMPARISON TABLE AND MODEL SELECTION

Model	MSE	R² (Train)	R² (Test)
Simple Linear Regression	1182893.25	0.75	0.75
Multiple Linear Regression	1182893.25	0.75	0.75
Random Forest Regression	497902.83	0.89	0.89
Ridge Linear Regression	1182893.32	0.75	0.74
Lasso Regression	1183106.69	0.76	0.75

min_samples_leaf	m-depth=	MSE	R2
= 1	3	1879582.69	0.60
	5	1270651.87	0.73
	10	542998.78	0.88
	15	500336.49	0.89
	20	499725.07	0 .89
	25	499725.07	0.89
= 5	3	1879582.69	0.60
	5	1270651.87	0.73
	10	542998.78	0.88
	15	500336.49	0.88
	20	499725.07	0 .89
	25	499725.07	0.89
= 10	3	1879582.68	0.60
	5	1270651.87	0.73
	10	542998.78	0.88
	15	500336.49	0.88
	20	499725.07	0.89
	25	499725.07	0.89



Streamlit



CONCLUSION

Project Overview:

- Aimed to analyze the synchronization between energy consumption and production across France.
- Focused on regional consumption behaviors and energy production landscape, including nuclear and renewable sectors.

Objectives and Insights:

- Analyzed regional energy consumption patterns and dissected the contributions of various energy sectors.
- Integrated temperature data to refine consumption forecasts, reflecting the impact of weather on energy demands.

Modeling Excellence:

- Explored statistical and machine learning models, with an optimized random forest model excelling in accuracy and efficiency.
- Provided robust national and regional consumption forecasts, essential for strategic energy sector decision-making.

Strategic Implications:

- Enhanced predictive strategies by integrating climate data, highlighting the interdependence between climate and energy consumption.
- Advocated for nuanced modeling approaches, balancing model complexity with performance to avoid overfitting.

Forward Momentum:

- Established a foundation for advanced, data-driven energy management strategies.
- The methodologies and insights pave the way for future research, contributing to energy sustainability and resilience amid global climate change.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to our mentor, Tarik Anouar and The DataScience Team for their unwavering patience, invaluable guidance, and continuous support throughout this project.

MARYAM MORADI
SARA CERRETO
LEENA WARUNKAR