Energy Behaviour Prediction of Prosumers

PURPLE - TEAM 4

Kelly, Peifen, Rish, Shalaka, Tony

I. INTRODUCTION

Business Problem

The primary business problem is addressing energy imbalance in the power grid caused by the increasing number of prosumers — consumers who both consume and generate energy. The goal is to create an energy prediction model for prosumers to reduce energy imbalance costs.

STAKEHOLDERS

- Energy Companies
- Prosumers
- Renewable Energy Advocates

OPPORTUNITIES

- Reduced Imbalance Costs
- Grid Stability
- Promoting Renewable Energy

CHALLENGES

- Unpredictable Behaviour
- Increasing Operational Costs
- Grid Instability

ETHICAL CONSIDERATIONS

- Privacy
- Transparency
- Equity

I. WHY THIS DATASET

- Novelty of Prosumer Behavior
  - eic_count
  - installed_capacity
  - is_business
  - is_consumption
- Sustainable Energy Practices
- Renewable Energy Integration

II. DATA SOURCES & PREPROCESSING

Overview of Data Sources

- Energy Production and Consumption Data
- Weather Data
- Energy Prices (Gas & Electricity)
- Installed Photovoltaic Capacity Records

Data Collection Period: September 1, 2021, to May 29, 2023 (636 days, approx. 1.74 years)

5 datasets, **51** variables, **138** columns, **2,018,352** rows

DATA PREPROCESSING STEPS

1. CLEAN DATA
2. QUALITY CHECK
3. CREATE NEW COLUMNS
4. DATA MERGE
5. PERFORM EDA

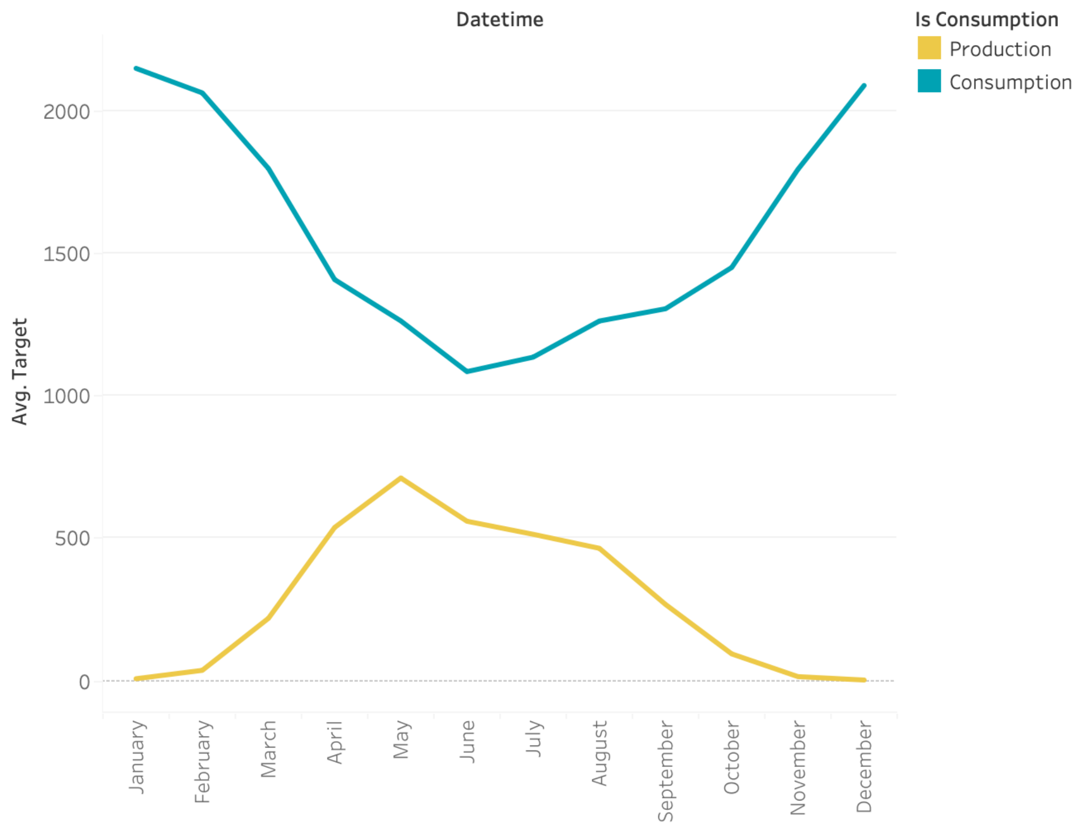
II. DATA SOURCES & PREPROCESSING

train.csv	gas_prices.csv	electricity_prices.csv	client.csv	historical_weather.csv
county is_business product_type target is_consumption datetime data_block_id row_id prediction_unit_id	origin_date forecast_date [lowest/highest] _price_per_mwh data_block_id	origin_date forecast_date euros_per_mwh data_block_id	product_type county eic_count installed_capacity is_business date data_block_id	datetime temperature dewpoint rain snowfall surface_pressure cloudcover_[low/mid/h igh/total] windspeed_10m winddirection_10m shortwave_radiation direct_solar_radiation diffuse_radiation [latitude/longitude] data_block_id

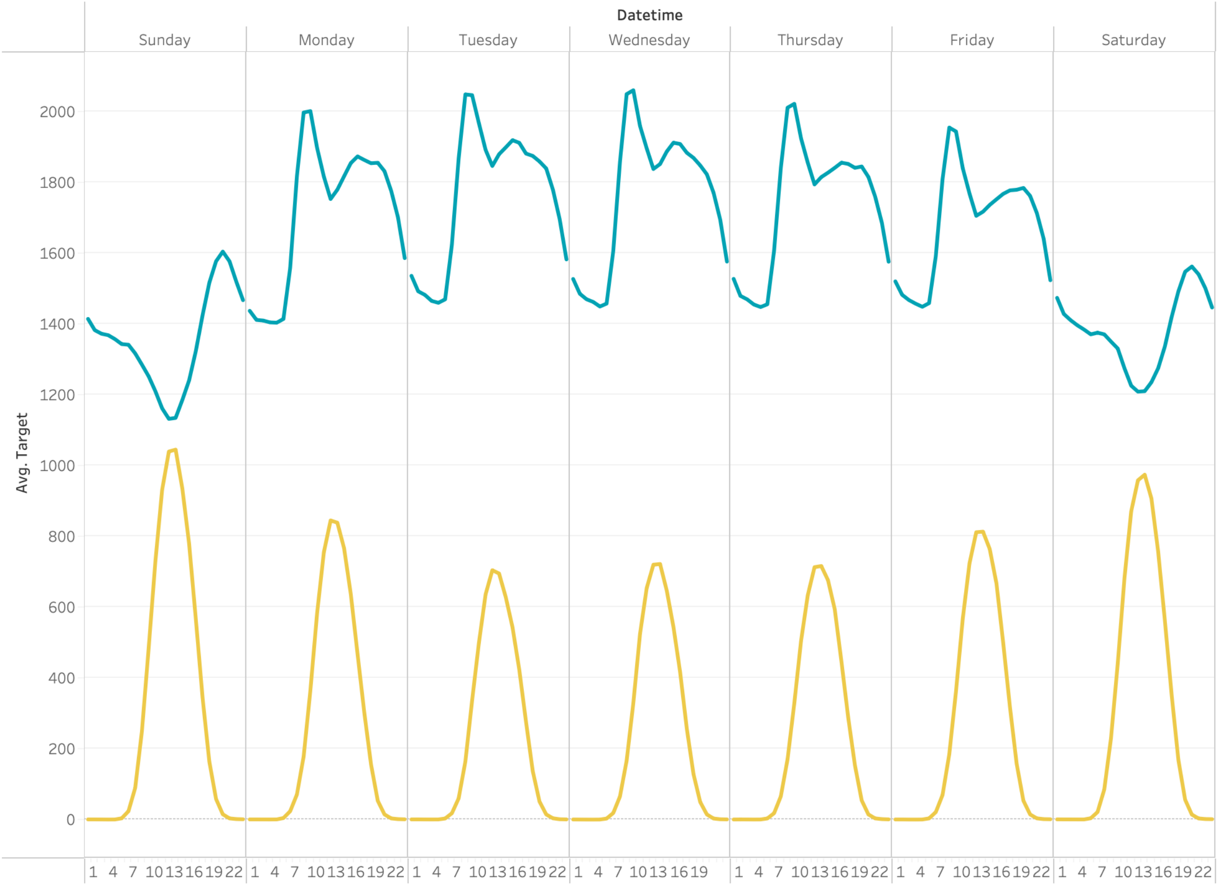
- *target - The consumption or production amount for the relevant segment for the hour. The segments are defined by the county, is_business, and product_type
- is_consumption - Boolean for whether or not this row's target is consumption or production.
- is_business - Boolean for whether or not the prosumer is a business.
- installed_capacity: Installed photovoltaic solar panel capacity in kilowatts

III. EXPLORATORY DATA ANALYSIS (EDA)

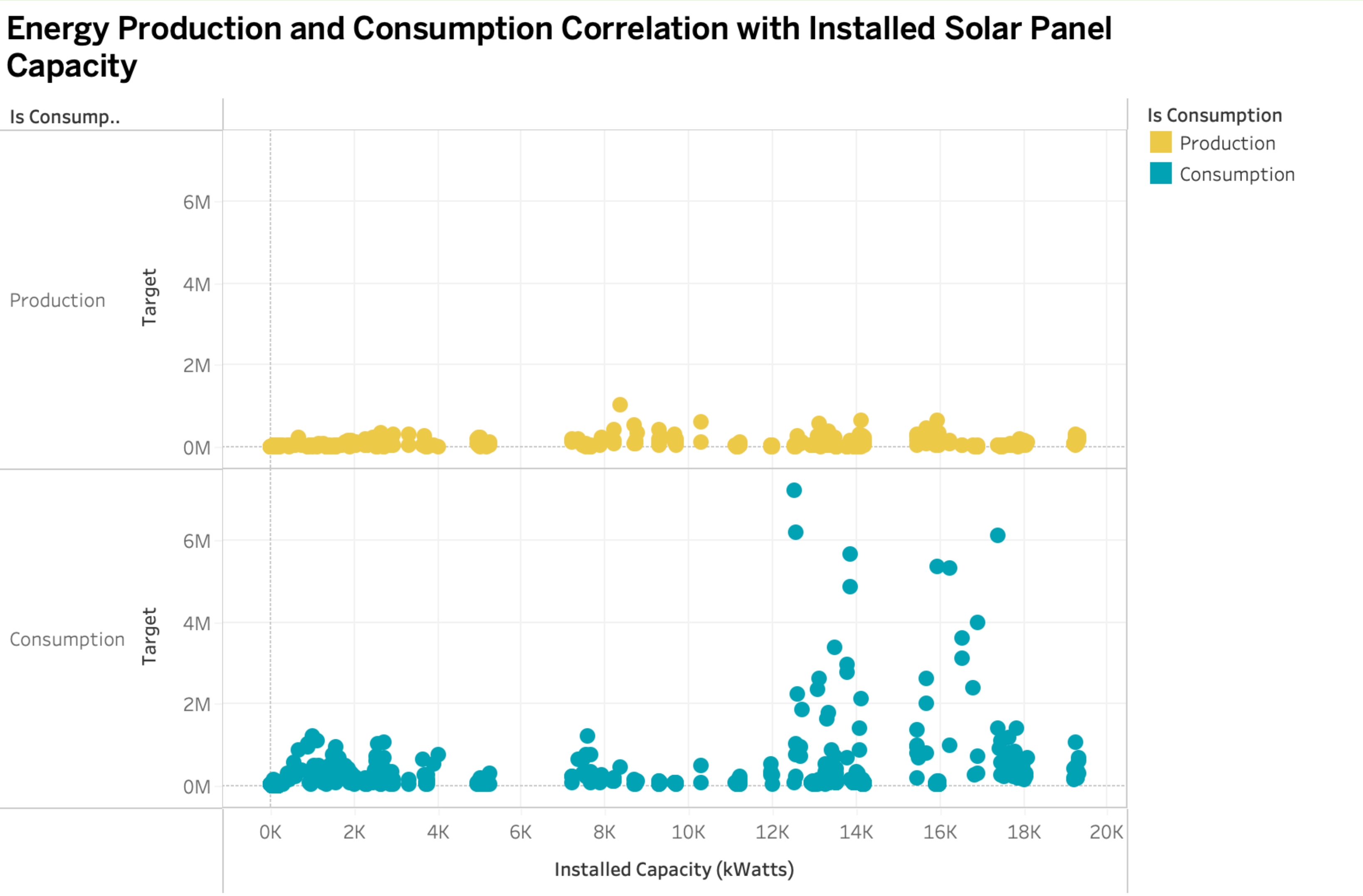
Temporal Features
Variation in consumption/production patterns based on the month of the year.

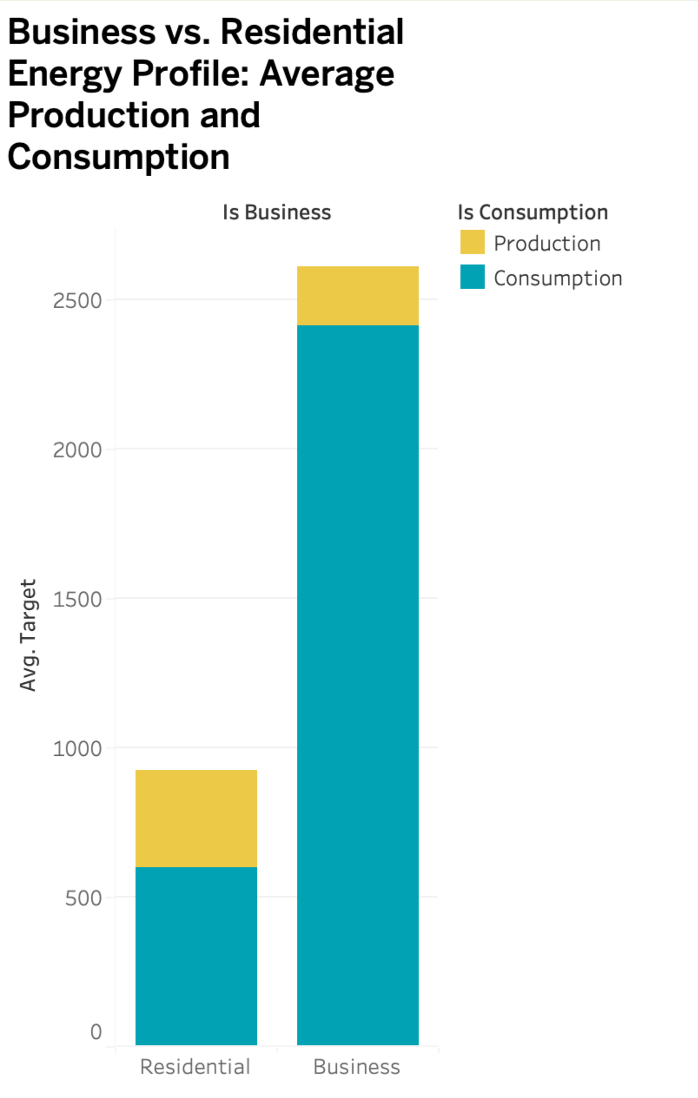


Temporal Features
Variation in consumption/production patterns based on the time of the day over a week

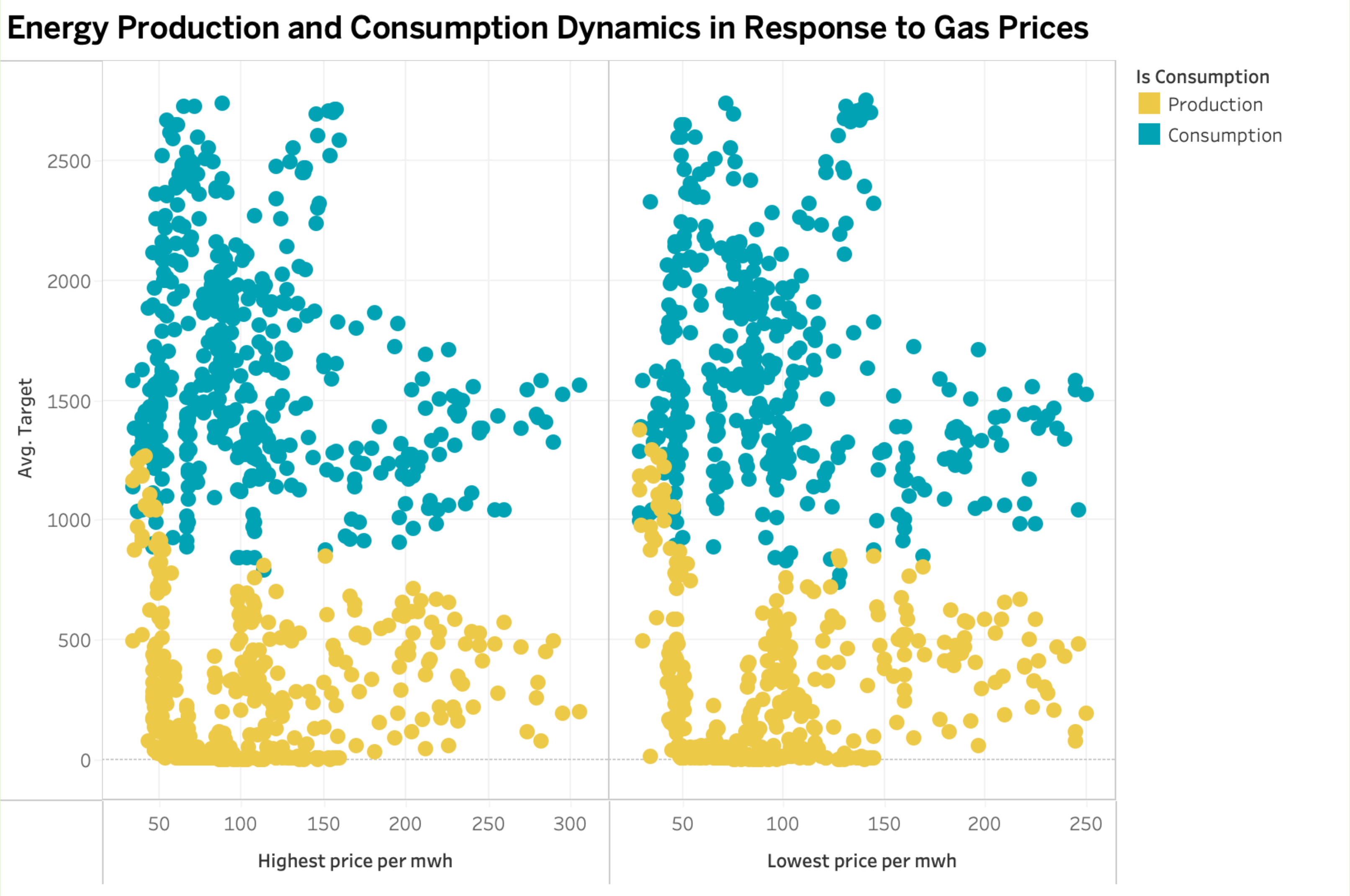


III. EXPLORATORY DATA ANALYSIS (EDA)

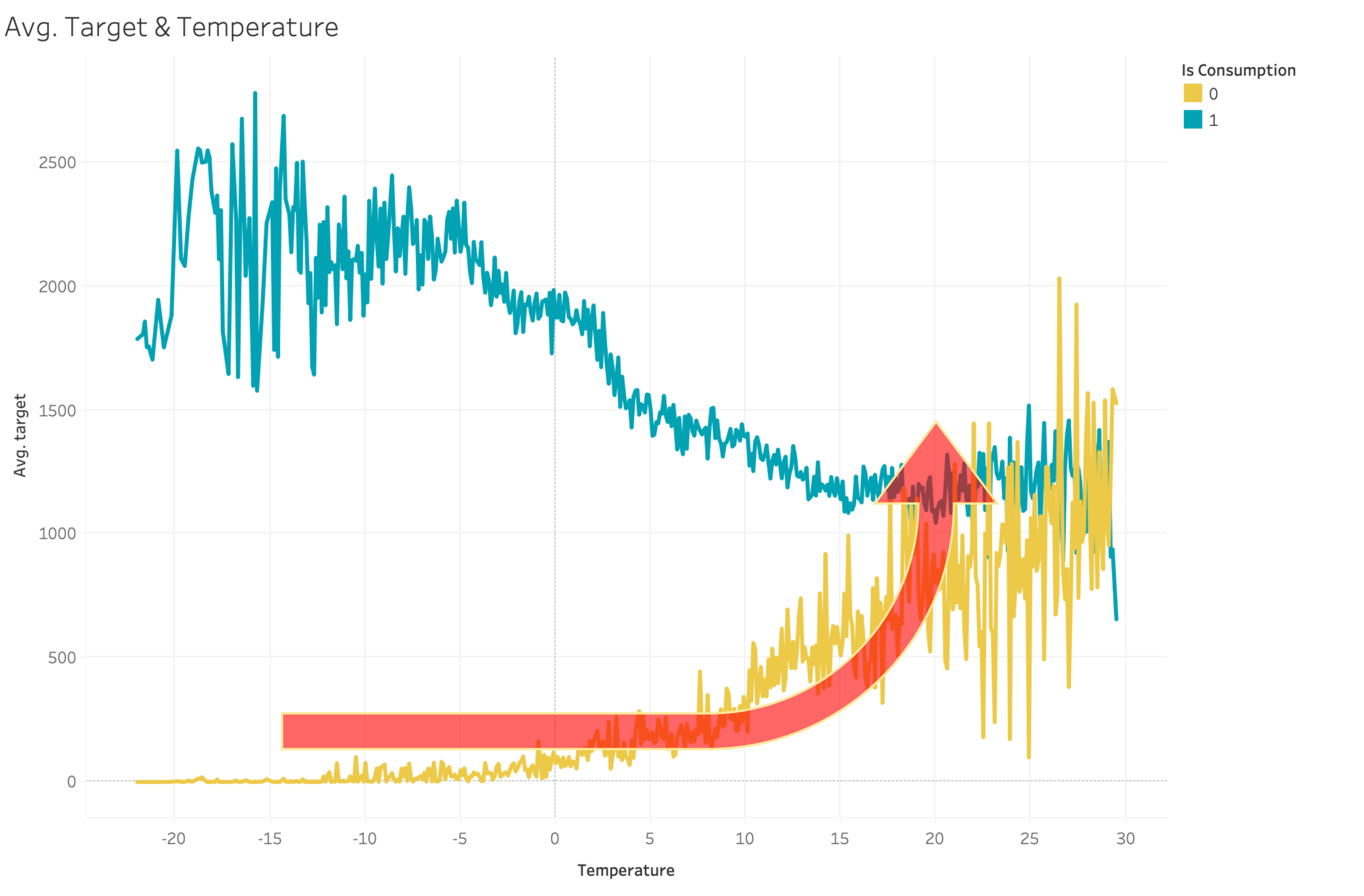




III. EXPLORATORY DATA ANALYSIS (EDA)



III. EXPLORATORY DATA ANALYSIS (EDA)



The trend of average of target for Temperature. Color shows details about Is Consumption.

III. EDA FINDINGS

- No linear relationship found between independent variables and the target variable.
- **Temperature-Related Insights:**
  - Identify temperature ranges with significant energy changes.
  - Adjust energy production based on weather forecasts, optimizing renewables during extreme temperatures.
- **Time-Related Insights (Hour of the Day):**
  - Identify peak hours for energy consumption, aiding utility planning.
  - Explore pricing structures to balance energy demand throughout the day.

IV. FEATURE SELECTION

```
'county, is_business, product_type, target, is_consumption, datetime, row_id, prediction_unit_id, date, installed_capacity, euros_per_mwh,
lowest_price_per_mwh, highest_price_per_mwh, temperature, dewpoint, rain, snowfall, surface_pressure, cloudcover_total, cloudcover_low, clo
udcover_mid, cloudcover_high, windspeed_10m, winddirection_10m, shortwave_radiation, direct_solar_radiation, diffuse_radiation, latitude, l
ongitude'
```

TOTAL FEATURES (23)

- Numerical features (20)
- Categorial Features (3): is_business, product_type, is_consumption

FEATURE ENGINEERING

- Categorical features: one-hot encoding
- Numerical features: normalization, convert datetime

V. MODEL SELECTION

REGRESSION MODELS

Model	Linear	Lasso/Ridge	Polynomial	MLP
	$y = ax + b$		$y = \mathrm{a}_1 x + \mathrm{a}_2 x^2 + \cdots + \mathrm{a}_\mathrm{k} x^k + b$	
Pros	• Simple • Interpretable	• Avoid overfitting	• Able to capture non-linear relationship	• Able to capture complex nonlinear relationships • Automatic feature learning
Cons	• Limited capability	• Limited capability	• Higher computational complexity	• Overfit

V. MODEL TRAINING

CV & HYPERPARAMETER TUNING

- Cross-validation

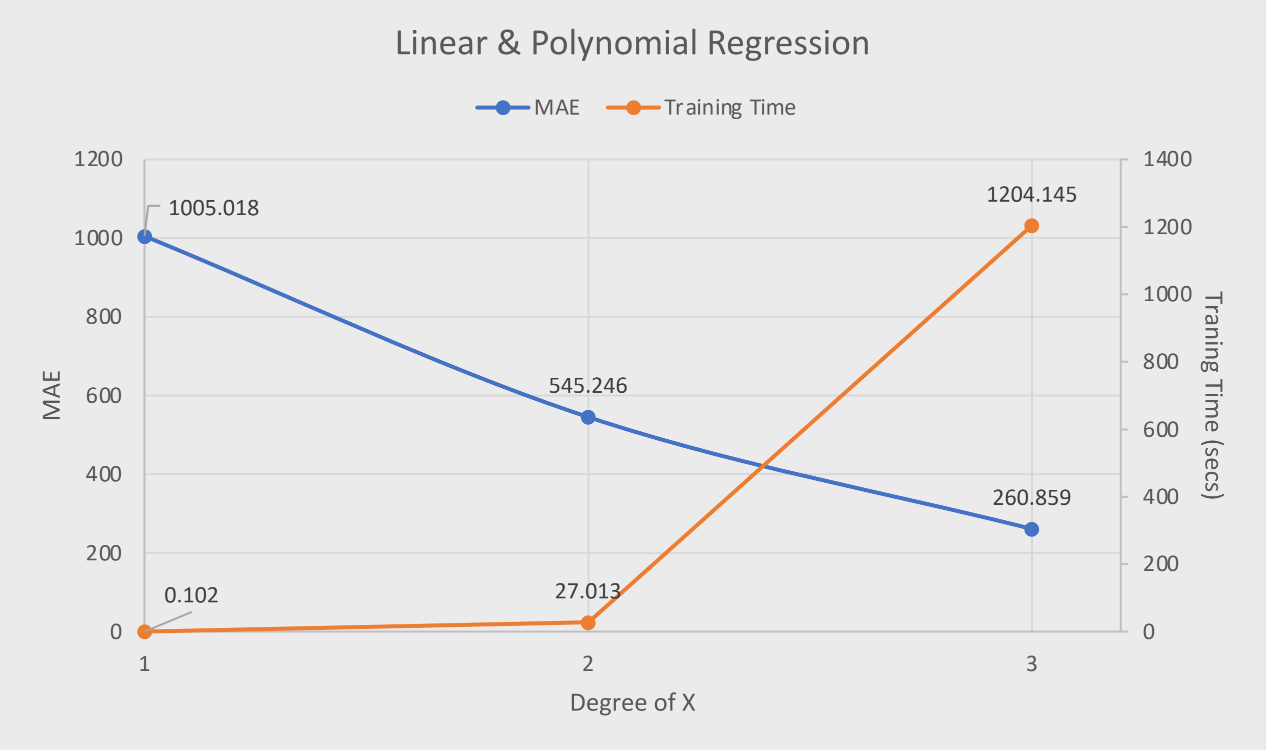
- Parameters to tune
  - Polynomial Regression: the degree of X
  - MLP: hidden units

VI. EVALUATION

Polynomial Regression vs MLP

Polynomial Regression

- Degree of X increases
  - **MAE** decreases
  - Training time increases
- Performance vs Cost



MLP

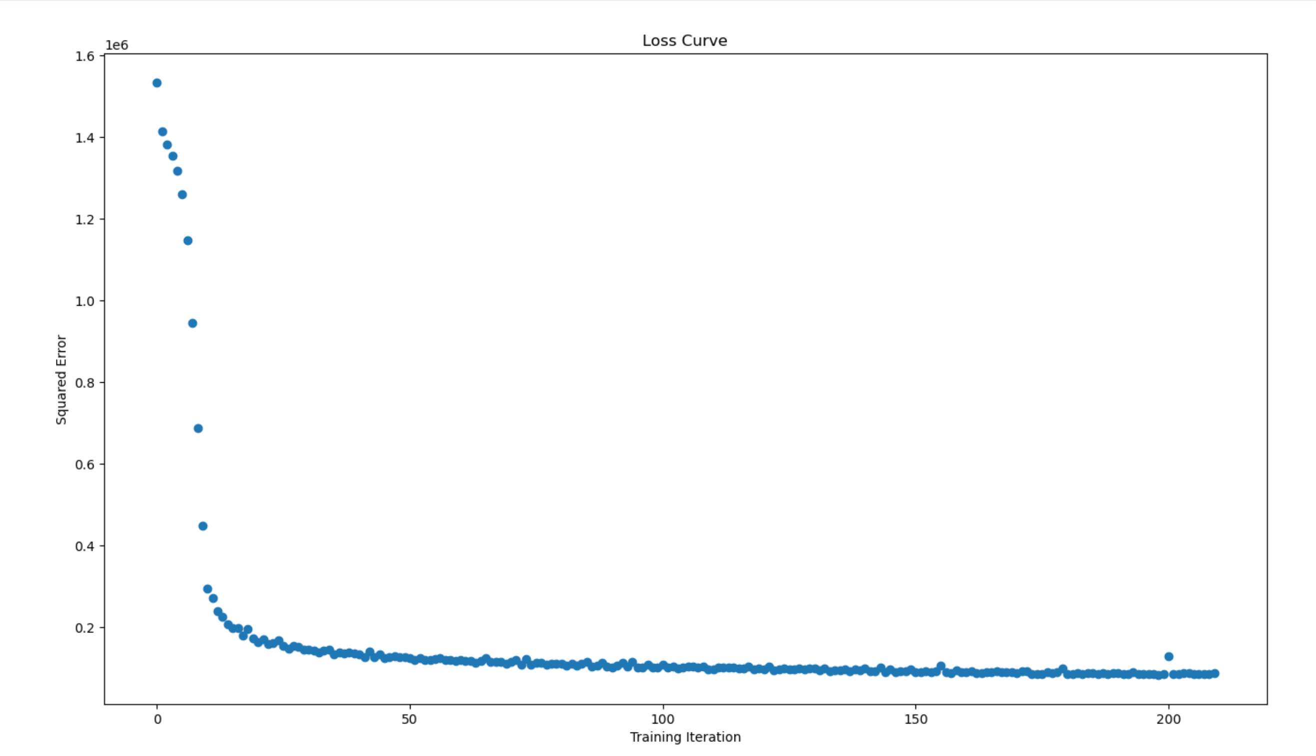
- Loss decreases rapidly in first few iterations

VI. EVALUATION

Polynomial Regression vs MLP

Polynomial Regression

- Degree of X increases
  - MAE decreases
  - Training time increases
- Performance vs Cost



MLP

- Loss decreases rapidly in first few iterations

VII. RESULT AND CONCLUSION

- Our analysis indicates that the Multilayer Perceptron (MLP) model provides the most accurate predictions for the target variable.

- While the MLP model excels in predictive accuracy, it inherently lacks interpretability. This poses a challenge in directly extracting insights from the model structure.

- Although the model's intricacies may not be directly interpretable, our intuition suggests that factors such as weather conditions, fuel prices, and electricity prices likely play pivotal roles in influencing energy demand. Further interpretative analysis could shed light on these aspects.

VIII. BUSINESS CONCLUSIONS & RECOMMENDATIONS

- Knowing how much energy prosumers use is crucial to avoid problems with power grid and to save money on logistics and operations.
- The Multilayer Perceptron model helps predict exactly how much energy prosumers will use.

IMPACT OF ACCURATE PREDICTION

- Reduction in Energy Imbalance Costs
- Enhanced Grid Stability
- Efficient Prosumer Integration
- Encouraging Prosumer Transition

STRATEGIC IMPLICATIONS:

- Cost savings and Efficiency
- Customer Friendly Systems

RECOMMENDATIONS –

- **Continuous Model Improvement:** Ongoing refinement of the predictive model as more data becomes available.
- **Collaboration with Experts:** to identify and prioritize the most influential factors affecting energy consumption.

IX. ACKNOWLEDGMENTS AND REFERENCES

- The data for this project was provided by Enefit – an energy company in Baltic region.

- Citation: Kristjn Eljand, Martin Laid, Jean-Baptiste Scellier, Sohier Dane, Maggie Demkin, Addison Howard. (2023). Enefit - Predict Energy Behavior of Prosumers. Kaggle. https://kaggle.com/competitions/predict-energy-behavior-of-prosumers

THANKYOU