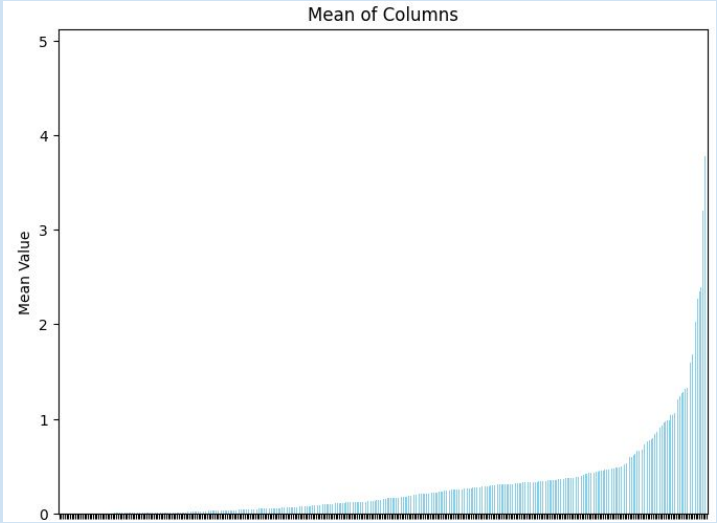
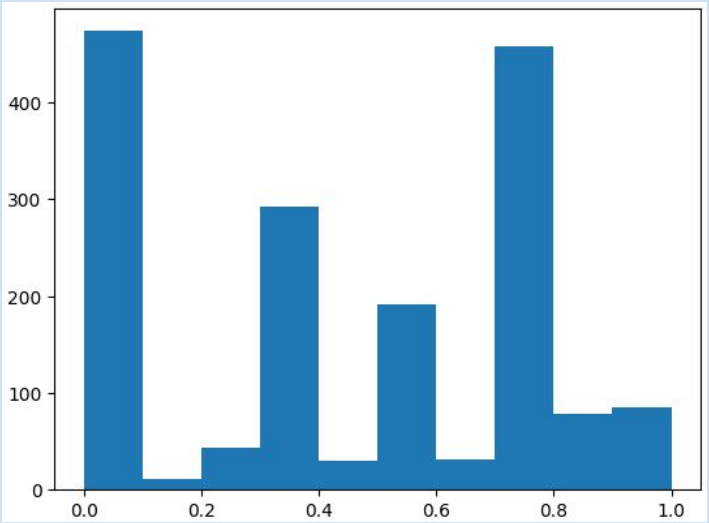
Modelling Tomorrow's Energy Needs

Team: Totoro

Data Features – Italy

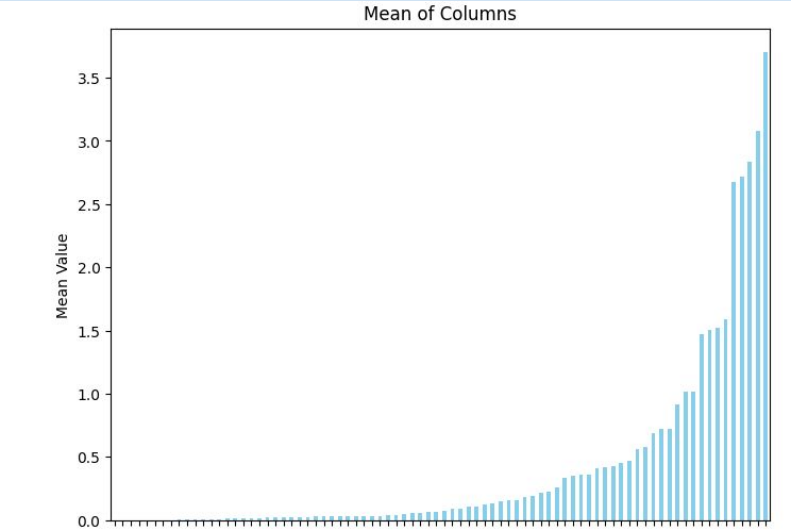


Mean usage per client

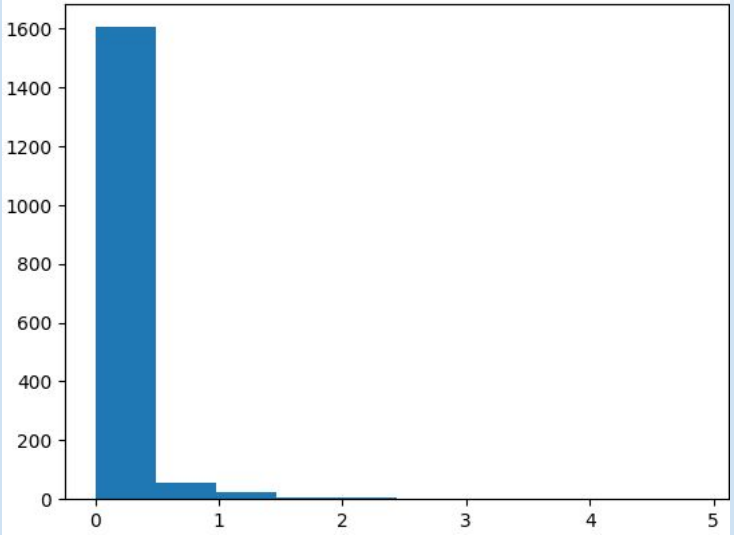


Distribution of column sparsity

Data Features – Spain

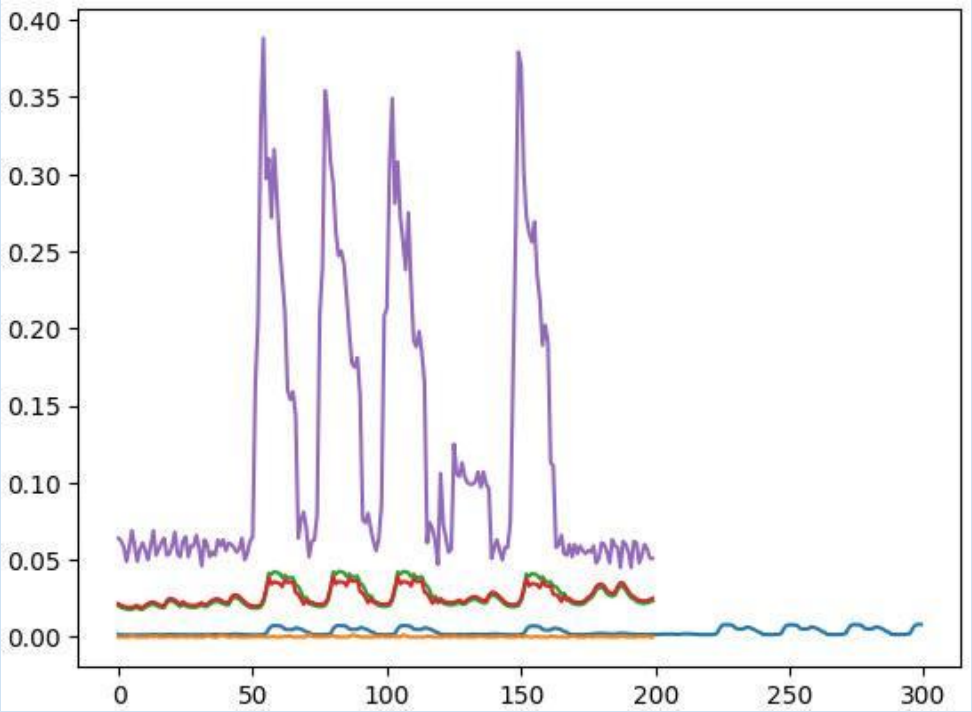


Mean usage per client



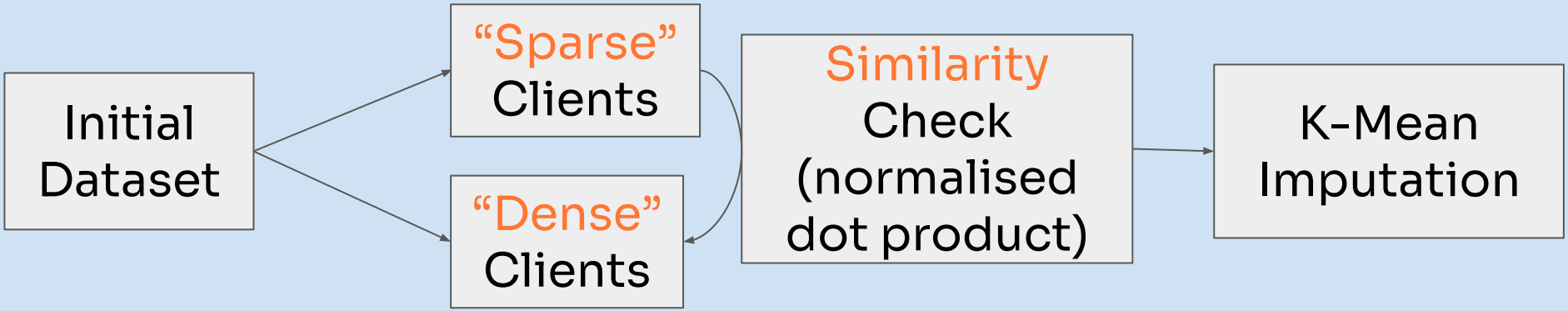
Distribution of column sparsity

Data Features – Mean and Standard Deviation



Imputation Techniques

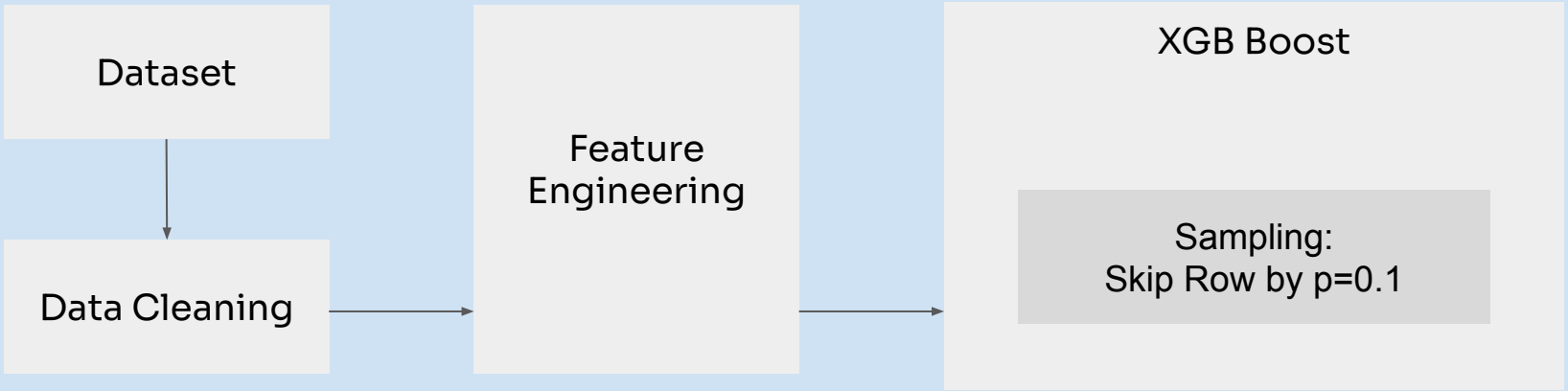
- Basic Techniques : Mean Imputation, KNN Imputation
- Customised Hot Deck Imputation:



Naive models

- **Simple Mean**
- **Linear Regression**
- **(S)ARIMA:** no stationarity, even by difference of predictions
- **Gaussian Processes**
- **Exponential Smoothing**

XGB Boost Approach



Feature Engineering

Hour Encoding:

$$\text{hour\_sin} = \sin\left(\frac{2\pi \cdot \text{hour}}{24}\right)$$

$$\text{hour\_cos} = \cos\left(\frac{2\pi \cdot \text{hour}}{24}\right)$$

Day of the Week:

$$\text{day\_of\_week\_sin} = \sin\left(\frac{2\pi \cdot \text{day\_of\_week}}{7}\right)$$

$$\text{day\_of\_week\_cos} = \cos\left(\frac{2\pi \cdot \text{day\_of\_week}}{7}\right)$$

Temperature:

$$\text{TMP\_squared} = \text{TMP}^2$$

Heating:

$$\text{heating\_degree} = \max(18 - \text{TMP}, 0)$$

$$\text{cooling\_degree} = \max(\text{TMP} - 18, 0)$$

Season, Weekend, Business Hours:

$$\text{season} = \left\lfloor \frac{(\text{month} \bmod 12) + 3}{3} \right\rfloor$$

$$\text{is\_weekend} = \begin{cases} 1, & \text{if day\_of\_week} \in 5, 6 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{is\_business\_hours} = \begin{cases} 1, & \text{if } 9 \leq \text{hour} \leq 17 \text{ and is\_weekend} = 0 \\ 0, & \text{otherwise} \end{cases}$$

XGB Results

~2M

Baseline score using simple mean

~299K

Score using XGB Boost

~230K

Score using LigthGBM

Thanks for the attention!

Questions?