Ensemble Learning Project

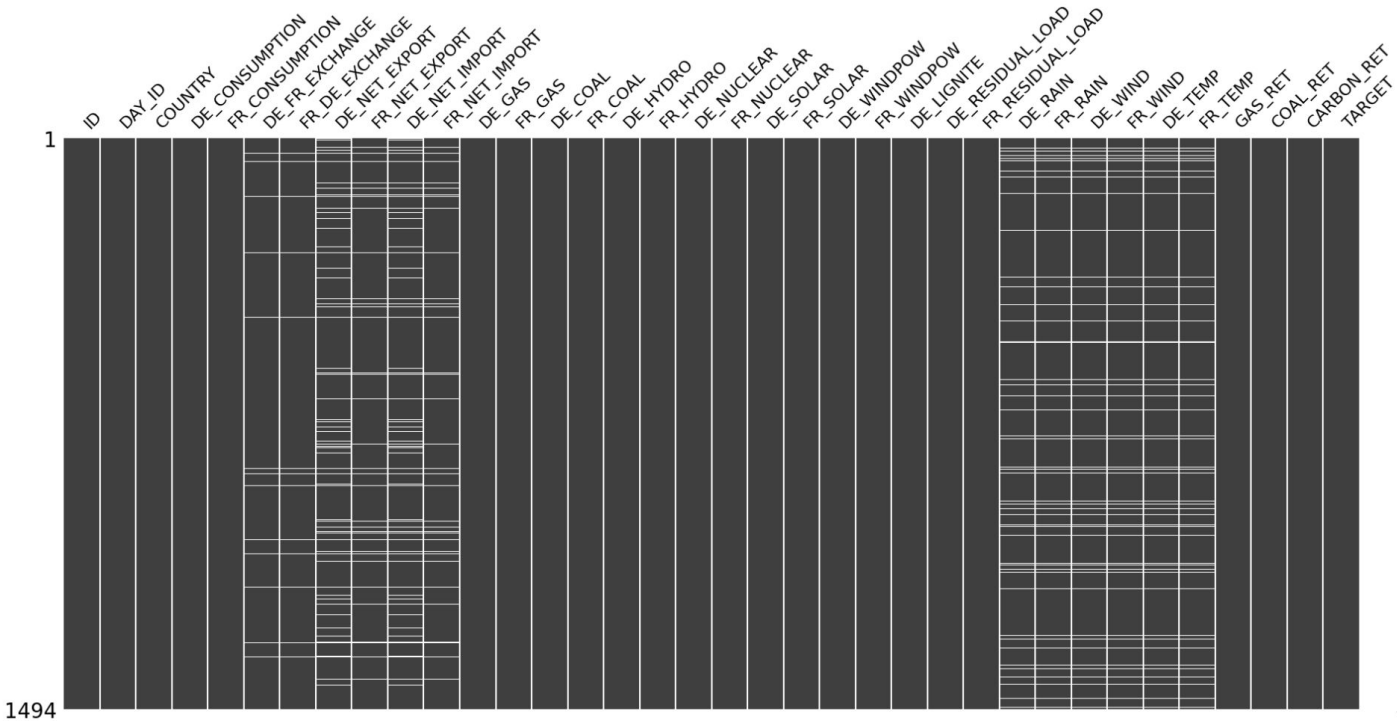
G3-T7

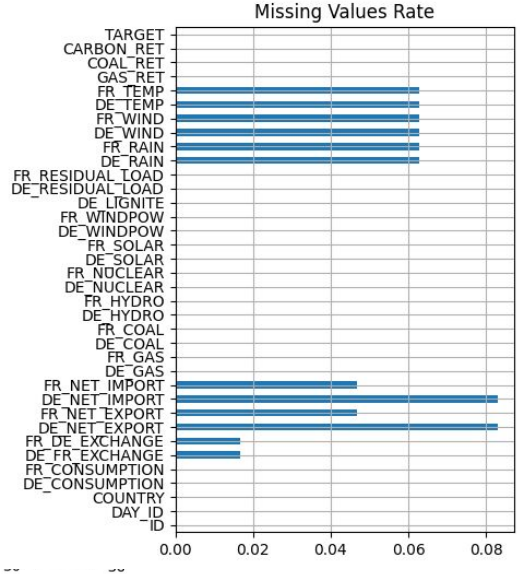
Team Members:

Irene Sunny, Wenjing Zhao

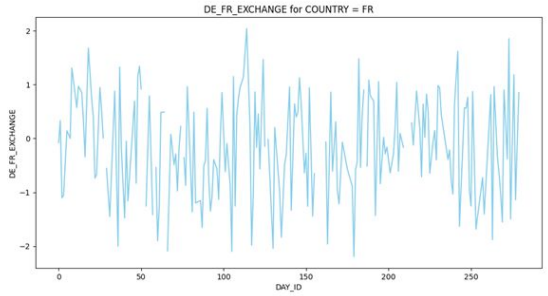
Hongyang Ye, Zheng Wan

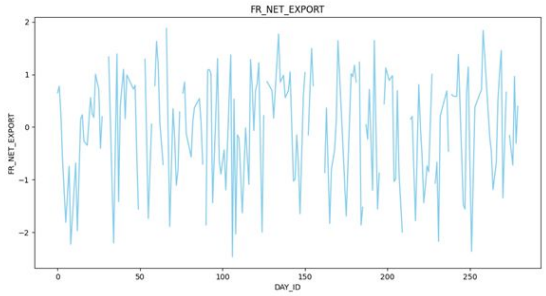
Missing values in features

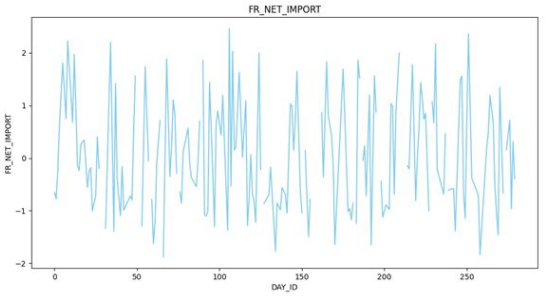


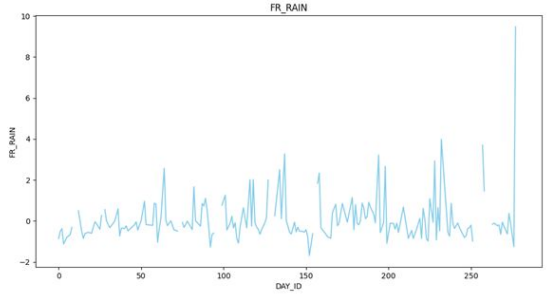


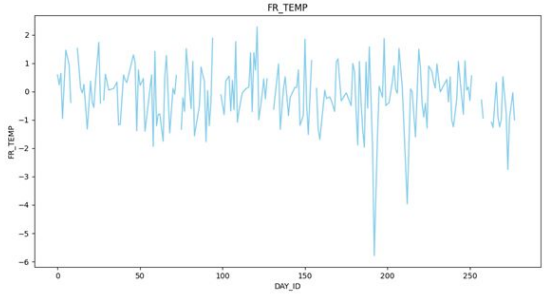
Missing values in features

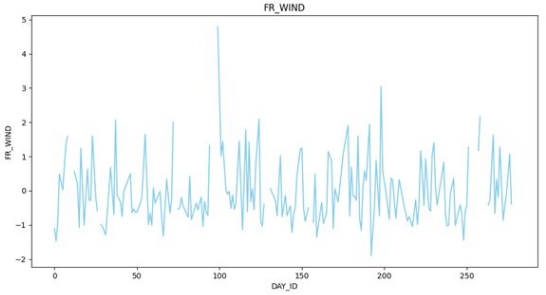






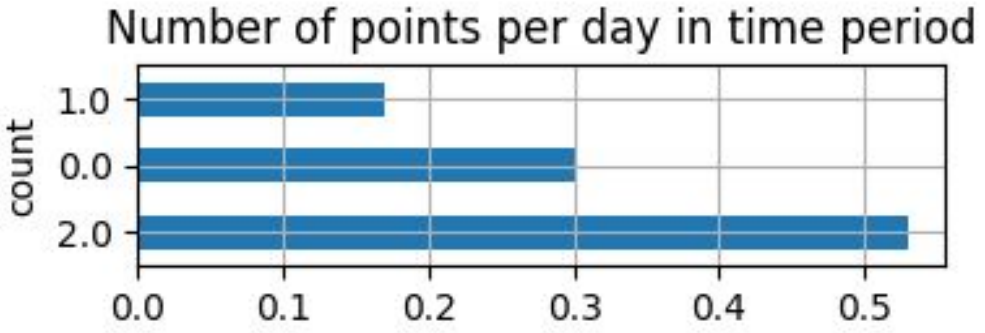




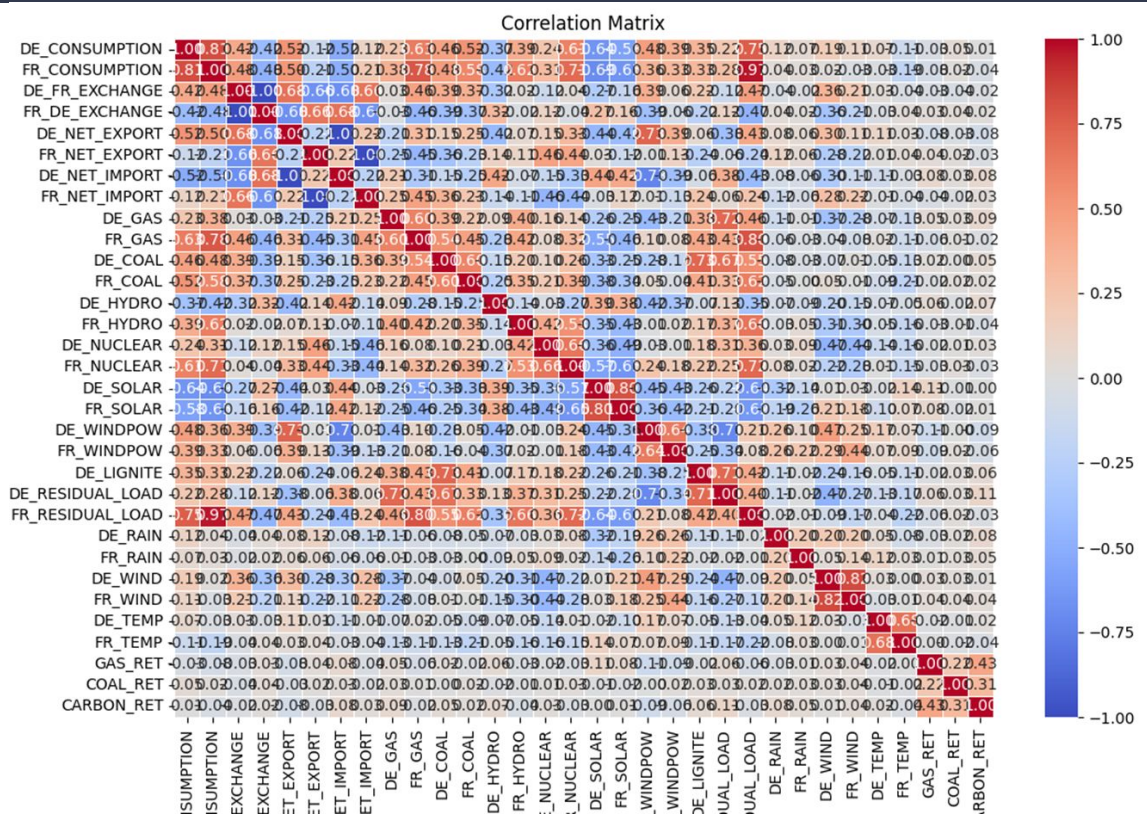


Missing values in sequential data (DAY_ID)

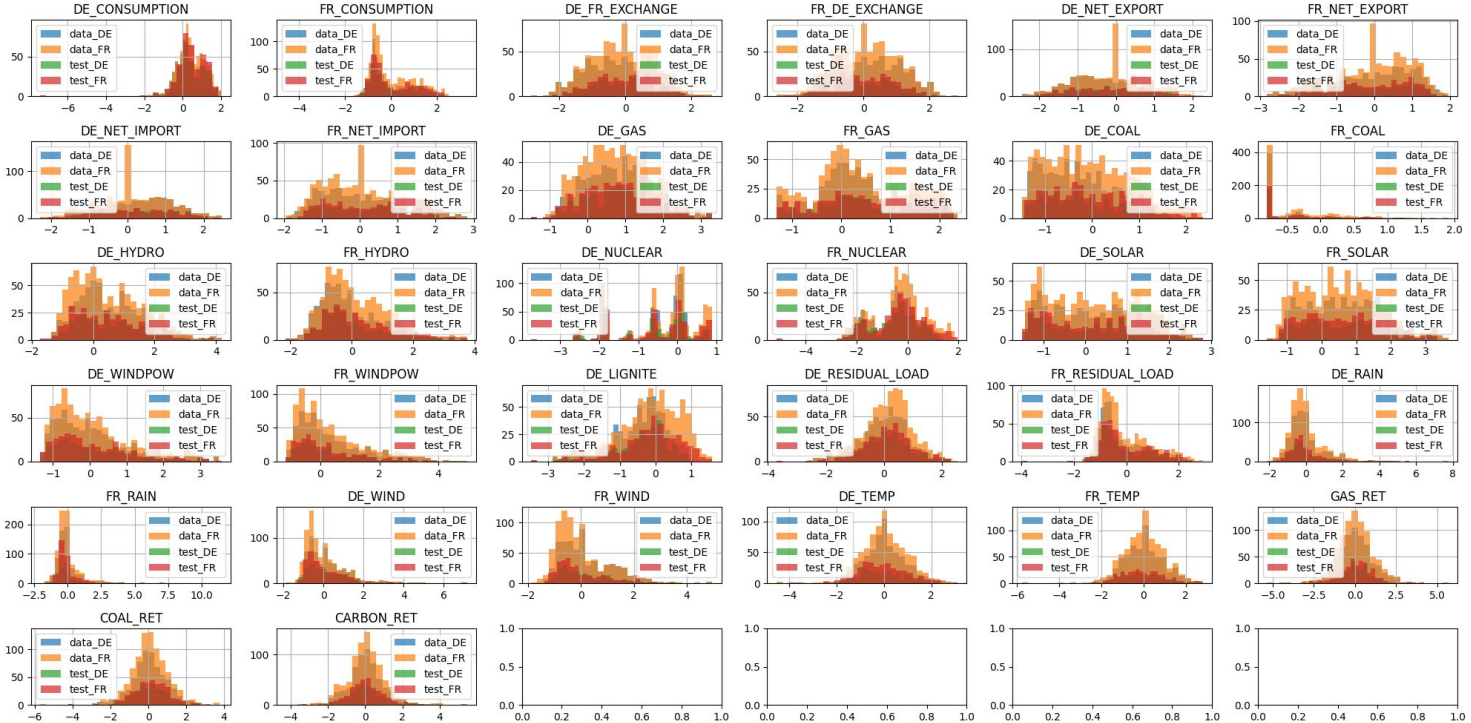
```
# For 208 days we have only information
of FR,no DE

# The occurrences of 'DE' and 'FR' in the
'COUNTRY' column.

(643, 851)
```



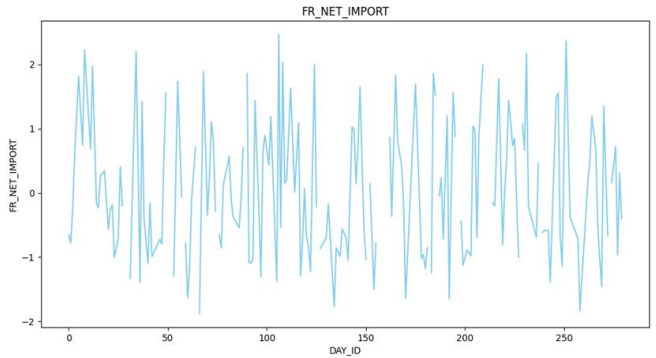
High correlations in features

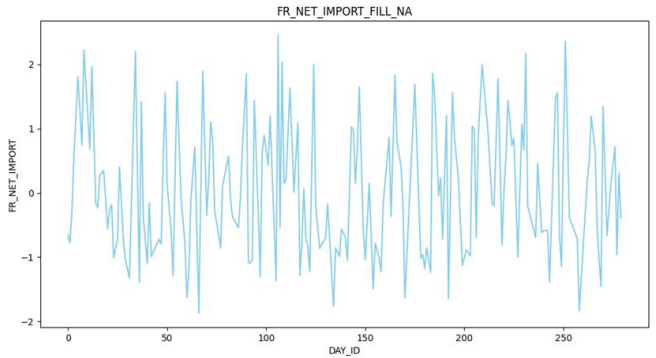


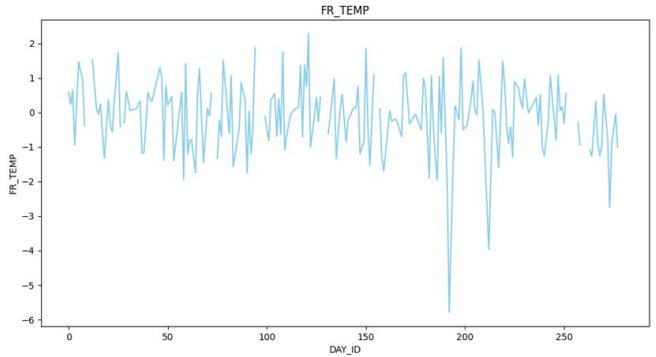
Feature Distribution (Normalization?)

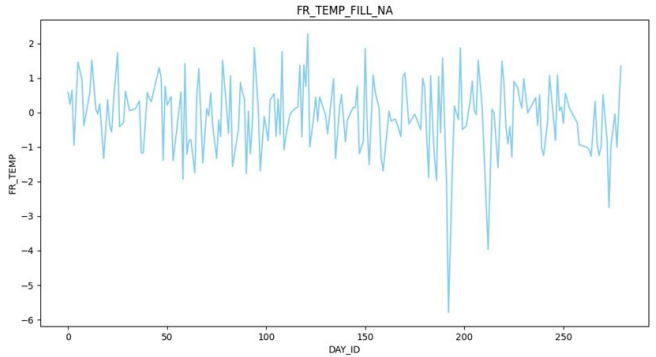


Solution – Missing values

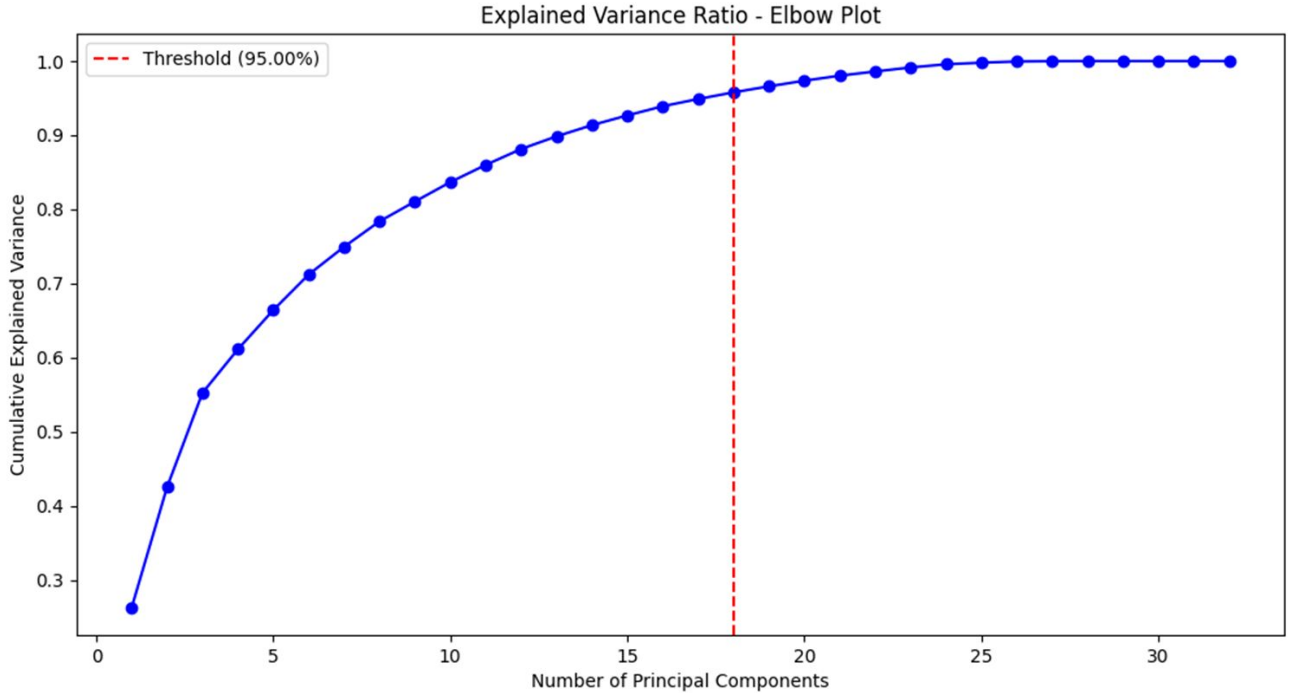




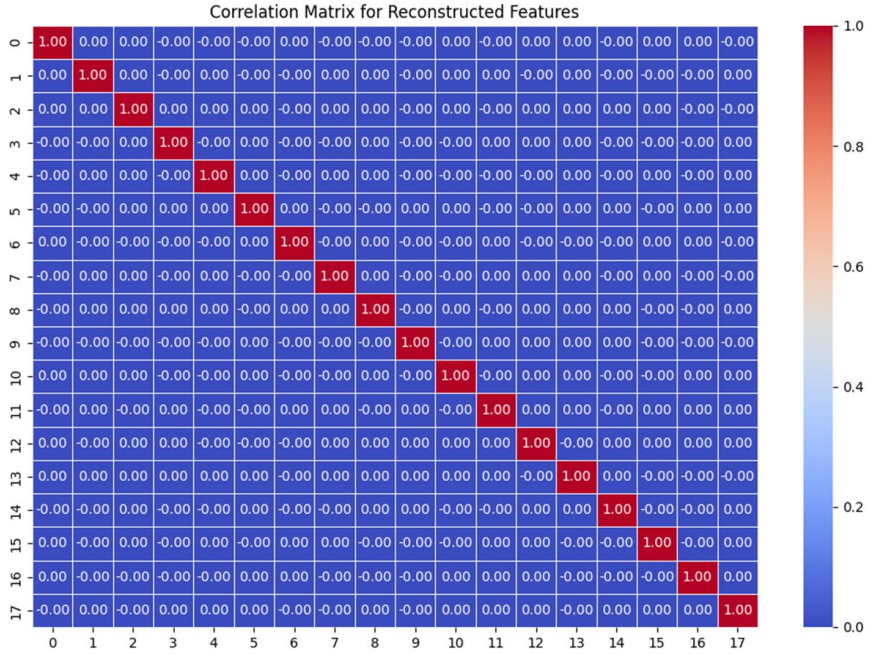




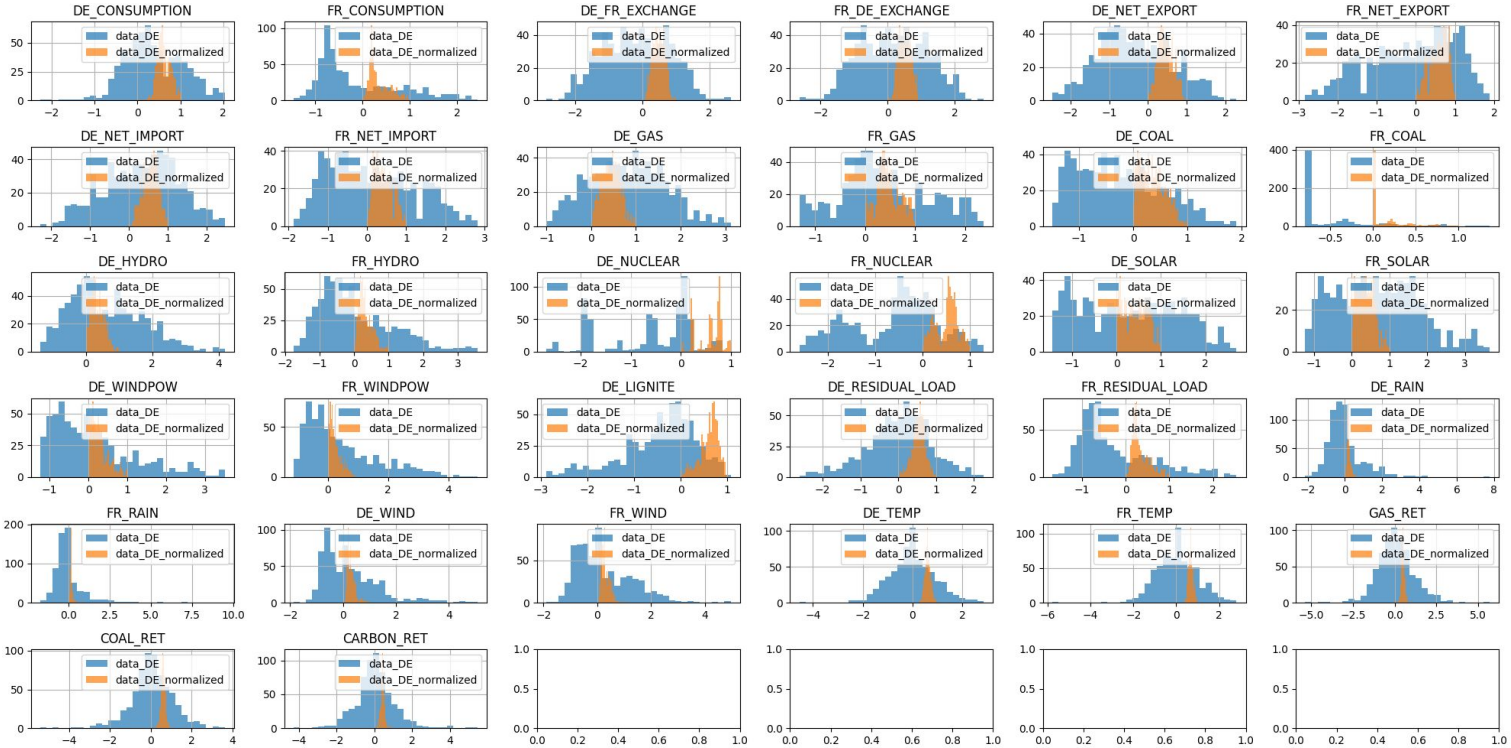
Solution - Multicollinearity



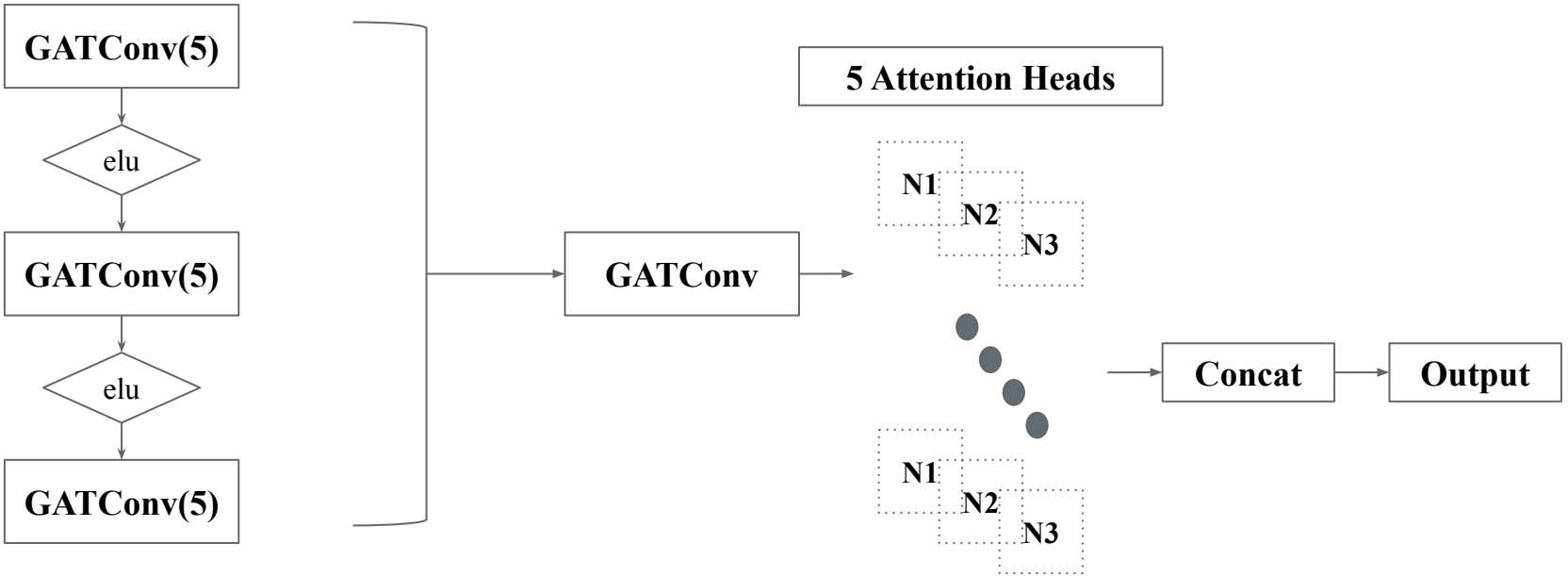
Solution - Multicollinearity



Solution - Normalization



Final Diagram



Summary

Observations

- There are missing values in some features
- There are missing values in days
- There are features that might be highly correlated
- We can’t tell if the data is already normalized or not

Solutions

- Replace missing values with appropriate approaches
- Check correlations between features
- Implement PCA on dataset to eliminate the effect of multicollinearity
- Conduct normalization on features