CIND 820 Capstone Project

Predicting Energy

Demand in Spain

By: Muhammad Zaka Shaheryar

Ryerson University

Introduction

- Tackling climate change with machine learning is the motivation behind this project

- For this purpose, project aim is to forecast energy demand accurately

- This will help in advance planning and dispatching of resources which in turn save environmental cost

- There most important entities in energy market are Transmission System Operator (TSO), power plant, commercial, and residential users

Research Questions

1. Which regression technique will accurately forecast the daily energy consumption demand using hourly period?

2. How to accurately forecast energy demand 24 hour in advance compared to TSO?

3. Using Classification, how to accurately forecast daily energy demand?

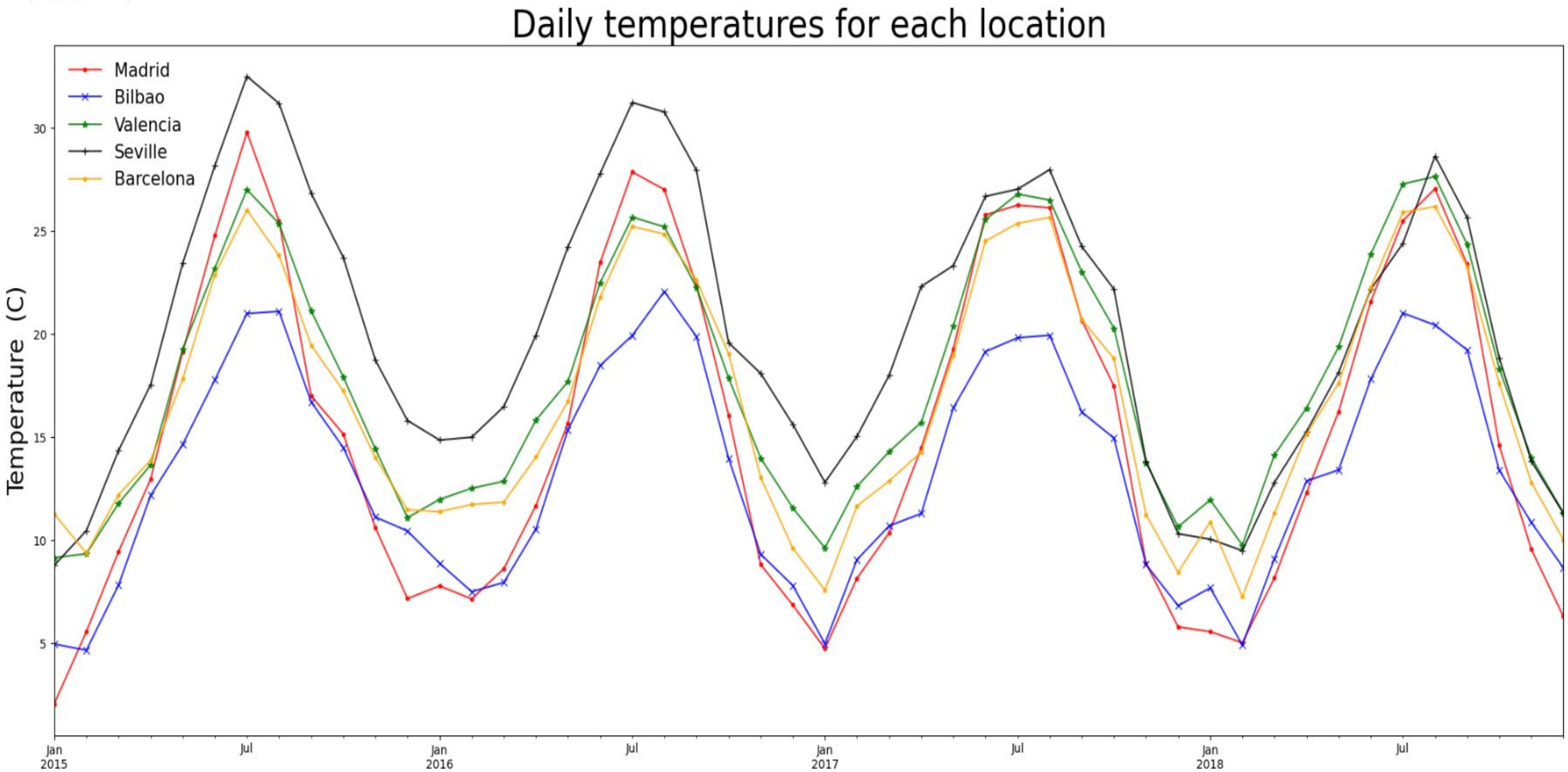
Ryerson University

Agenda

- Initial Analysis
- Exploratory Data Analysis
- Dimensionality Reduction
- Experimental Design
- Modelling
- Evaluation
- Improving the Model
- Conclusion

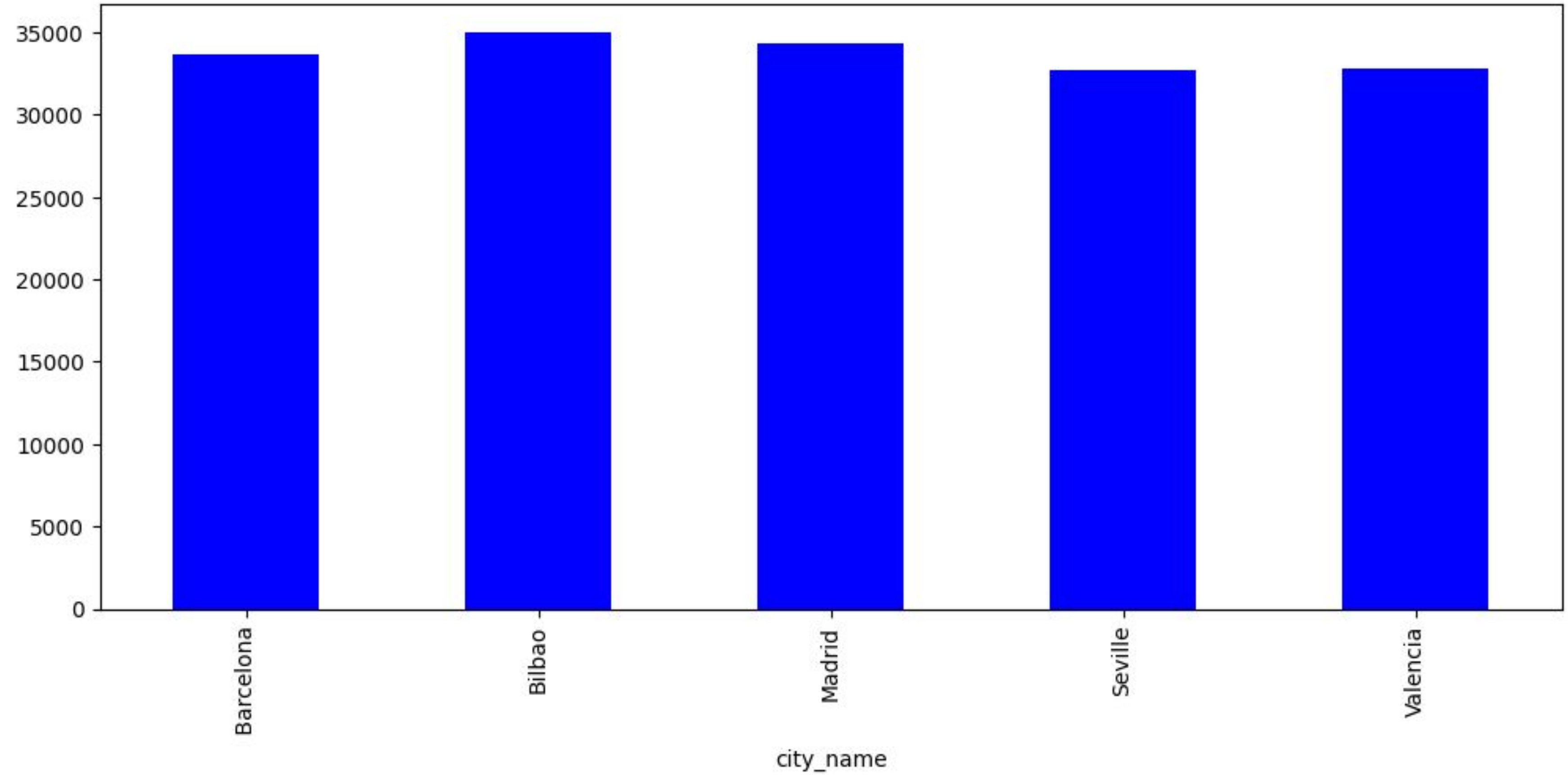
Trends and Patterns

1. Temperature Profile



Trends and Patterns Continued

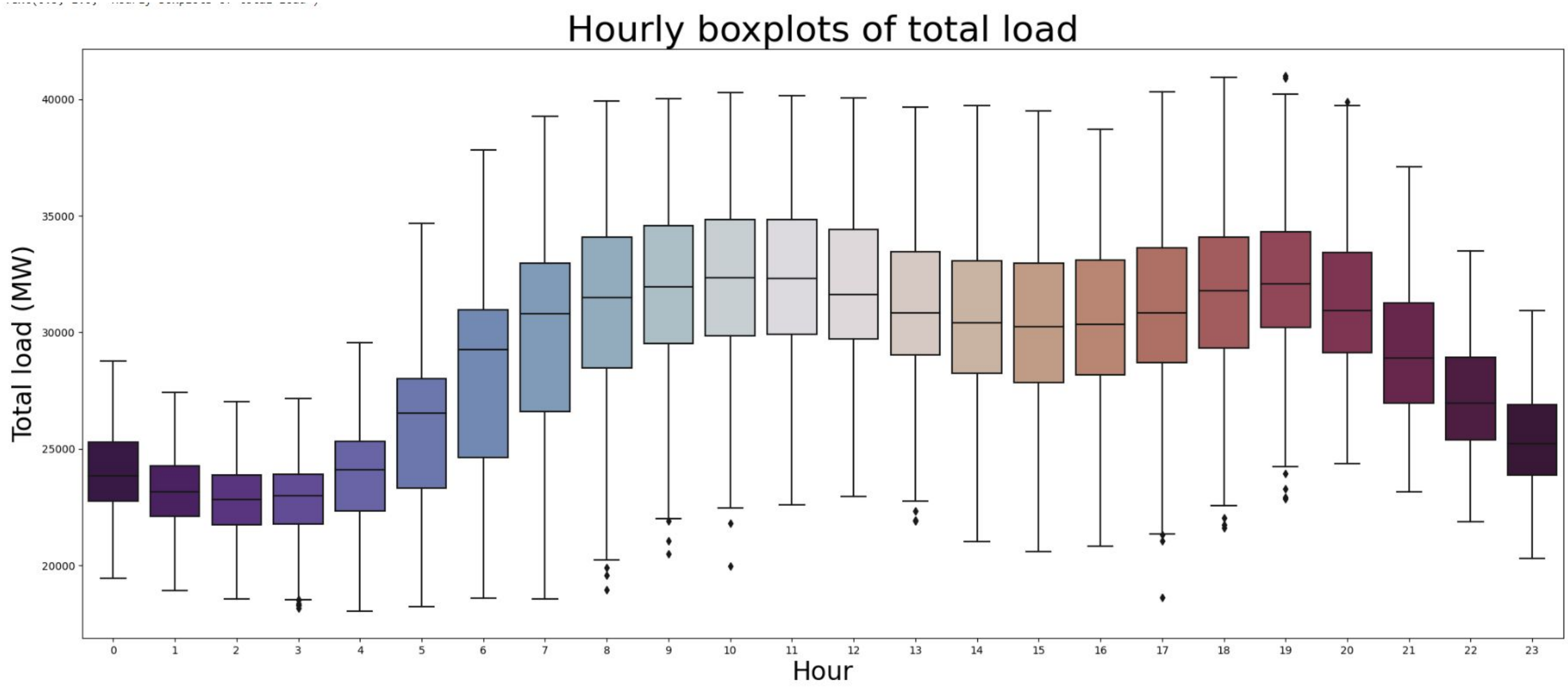
2. Number of Observations for each cities



Ryerson University

Trends and Patterns Continued

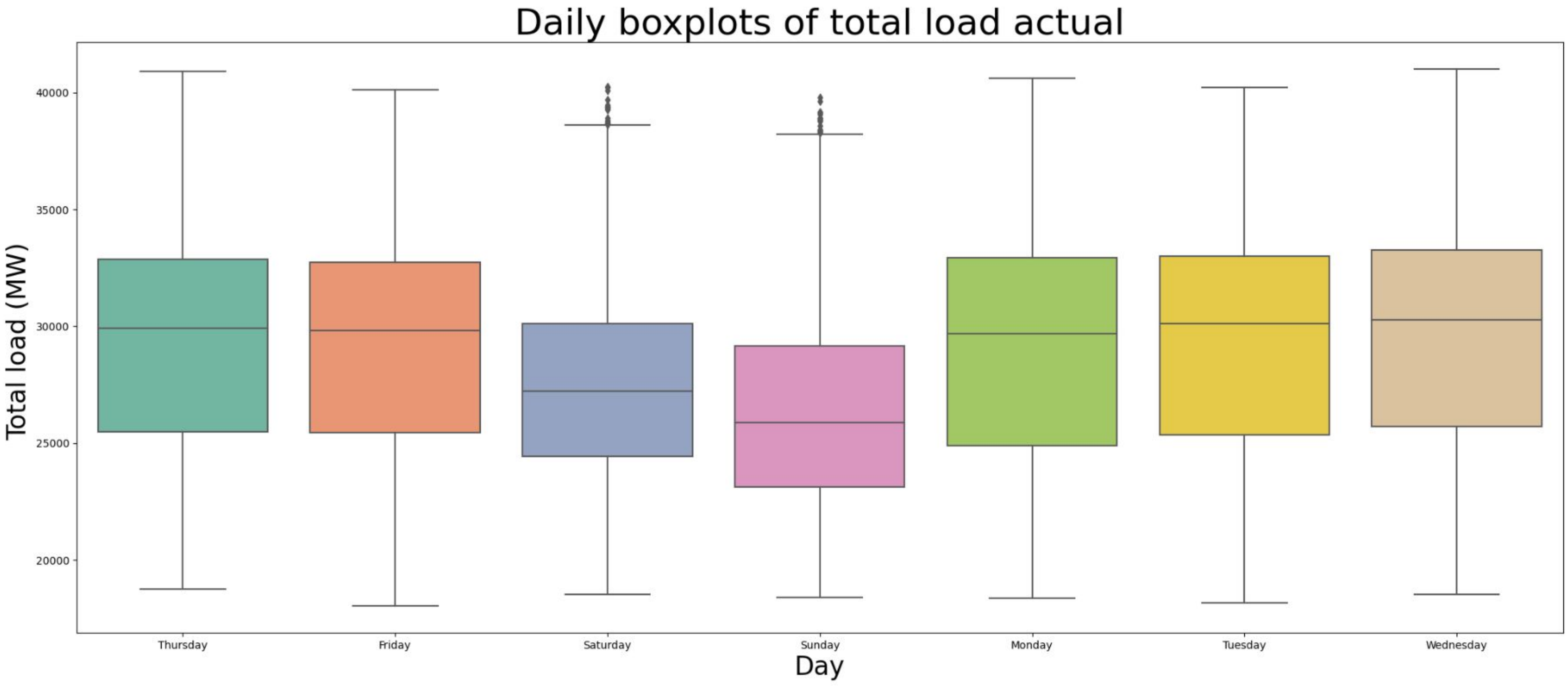
3. Hourly Load Profile



Ryerson University

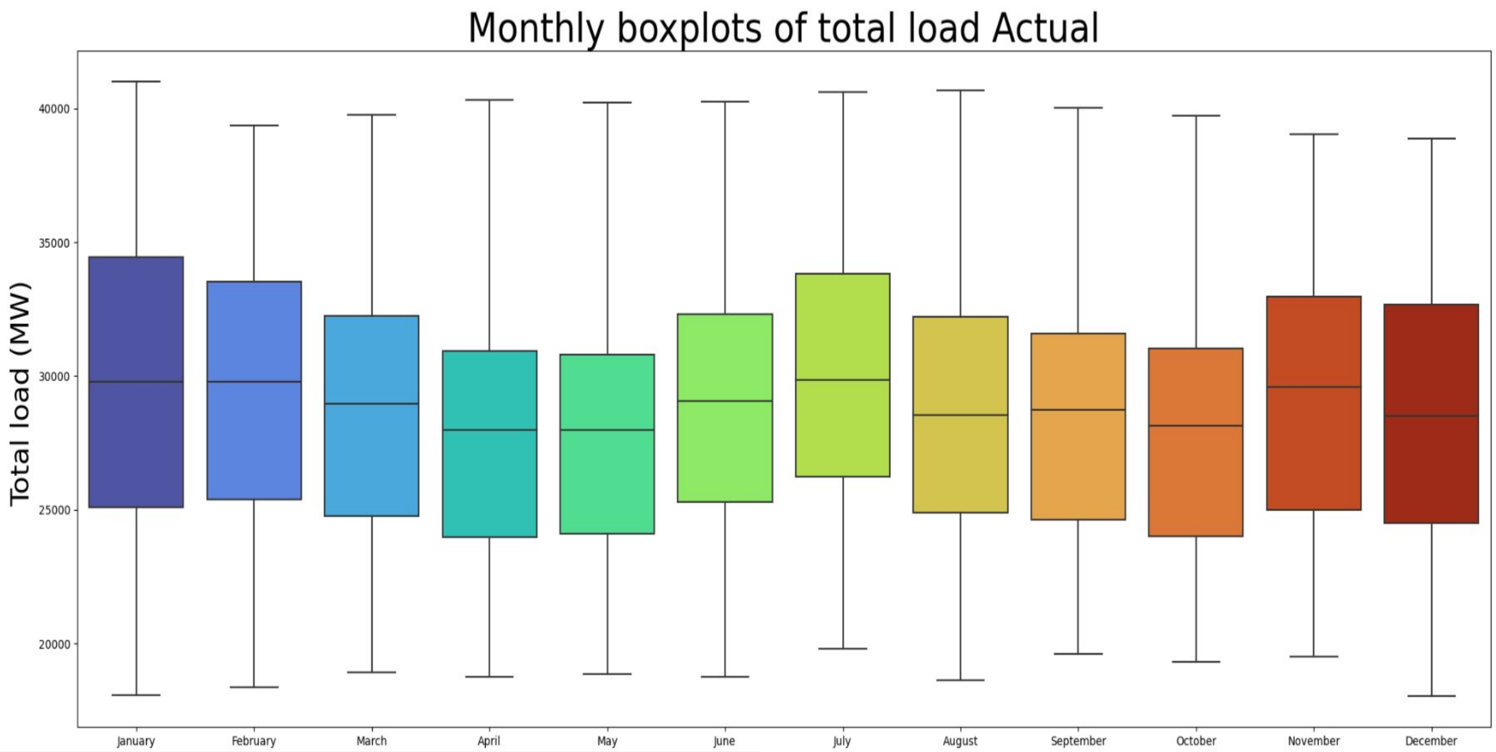
Trends and Patterns Continued

4. Daily Load Profile



Trends and Patterns Continued

5. Monthly Load Profile



Ryerson University

Literature Review

- Regression/Classification models considered are linear regression, KNN regression, Regression trees, Random forest regression, ANN, Light Gradient Boosted Machine, Support Vector Regression, Long-Short Memory Networks, ARX, NARX, AdaBoost, K-Neighbors Regressor

- Depending on case study, different models outperform other

- Performance of models is evaluated using metric such as R^2, RMSE, MAPE, NRMSE, ND for Regression analysis

- Performance of models is evaluated using metric such as accuracy, F1-score, Precision, and Recall for Classification analysis

- Input parameters considered can be categorized into time variable, weather variable, and electricity usage variable

Data Preparation

- Dataset is take from Kaggle and it consists of two big csv files
- First is energy_dataset file which contains 35064 rows and 29 columns
- Second is weather_features file which contains 178397 rows and 17 column
- The dataset has four years (2015-2018) of electrical consumption, generation, pricing, and weather data for five cities in Spain
- First mean is taken across all 5 cities in Spain to make the second file matchable with first file
- Then Both csv files are joined on common column for analysis
- Columns with more than 99% missing values or 99 % zero are dropped
- Redundant columns or categorical columns containing weather description are also dropped.

Data Preparation
Descriptive Statistics

Attribute	Distribution	Mean	Standard deviation
pressure (hPA)	**Not Normal**	**1073**	**6143**
humidity	**Normal (Left Skewed)**	**68**	**22**
wind_speed (km/h)	**Normal (Left Skewed)**	**2.5**	**2**
wind_deg (degrees)	**Not Normal**	**167**	**117**
rain_1h (%)	**Not Normal**	**0.08**	**0.4**
snow_3h (%)	**Not Normal**	**0.005**	**0.2**
clouds_all	**Not Normal**	**25**	**31**
temp_C (degree celsius)	**Normal**	**17**	**8**
temp_C_max (degree celsius)	**Normal**	**15**	**9**
temp_C_min (degree celsius)	**Normal**	**18**	**8**

Data Preparation
Descriptive Statistics continued

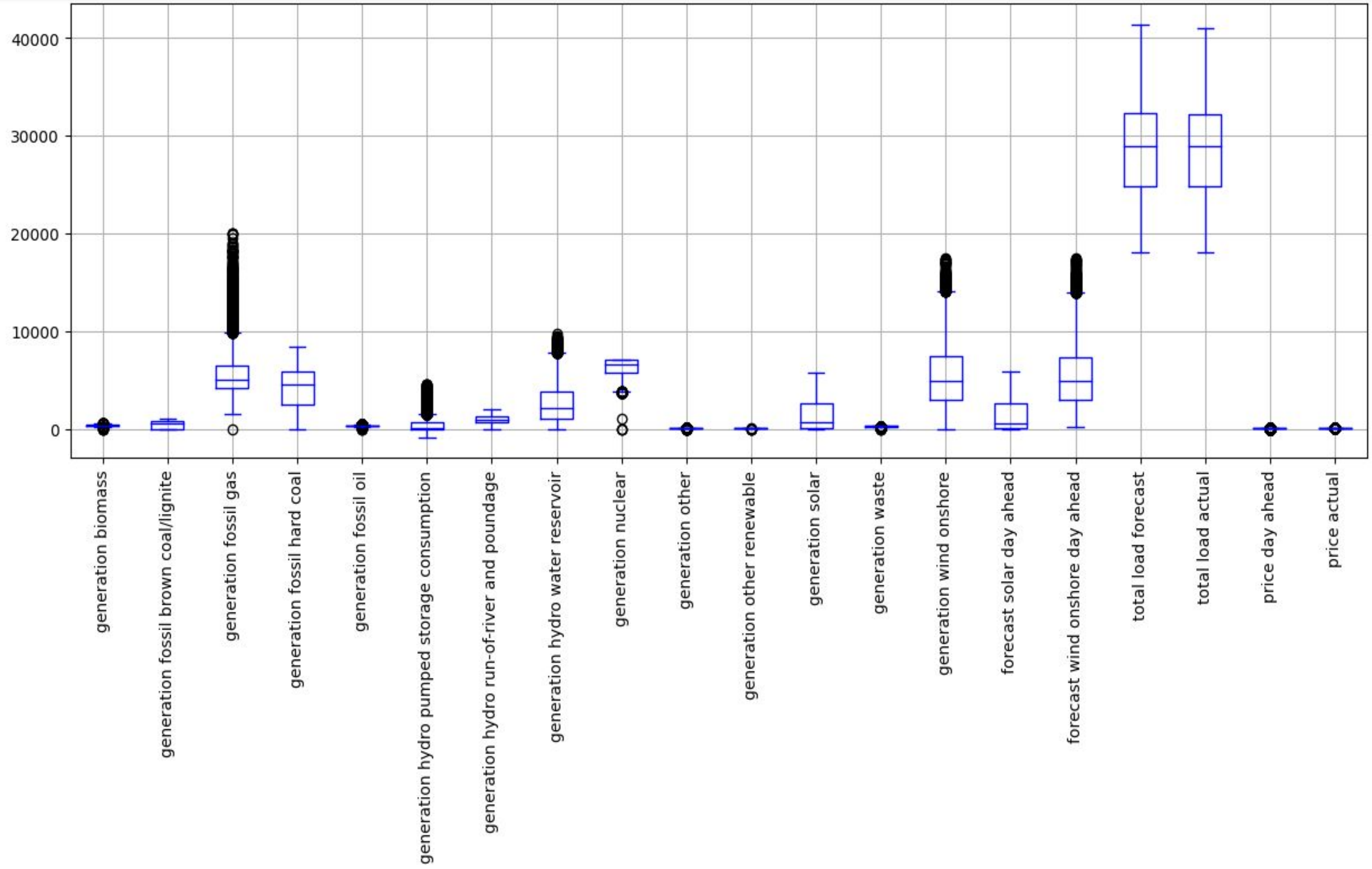
Attribute	Distribution	Mean	Standard deviation
generation biomass (MW)	**Hybrid**	**384**	**85**
generation fossil brown coal lignite (MW)	**Not Normal**	**448**	**355**
generation fossil gas (MW)	**Normal (Right Skewed)**	**5623**	**2202**
generation hard coal (MW)	**Normal**	**4256**	**1962**
generation fossil oil (MW)	**Normal**	**298**	**53**
generation hydro pumped storage consumption (MW)	**Not Normal**	**476**	**792**
generation hydro run of river and poundage (MW)	**Normal**	**972**	**401**
generation hydro water reservoir (MW)	**Normal (Right Skewed)**	**2605**	**1835**
generation nuclear (MW)	**Not Normal**	**6264**	**840**
generation other (MW)	**Not Normal**	**60**	**20**

Data Preparation
Descriptive Statistics continued

Attribute	Distribution	Mean	Standard deviation
generation other renewable (MW)	**Normal**	**86**	**14**
generation solar (MW)	**Not Normal**	**1433**	**1680**
generation waste (MW)	**Normal (Left Skewed)**	**269**	**50**
generation wind onshore (MW)	**Normal (Right Skewed)**	**5464**	**3214**
total generation (MW)	**Normal**	**27509**	**4105**
forecast solar day ahead (MW)	**Not Normal**	**1439**	**1678**
forecast wind onshore day ahead (MW)	**Normal (Right Skewed)**	**5471**	**3176**
total load forecast (MW)	**Normal**	**28712**	**4594**
total load actual (MW)	**Normal**	**28697**	**4575**
price day ahead (Euro)	**Normal (Left Skewed)**	**50**	**15**
price actual (Euro)	**Normal**	**58**	**14**

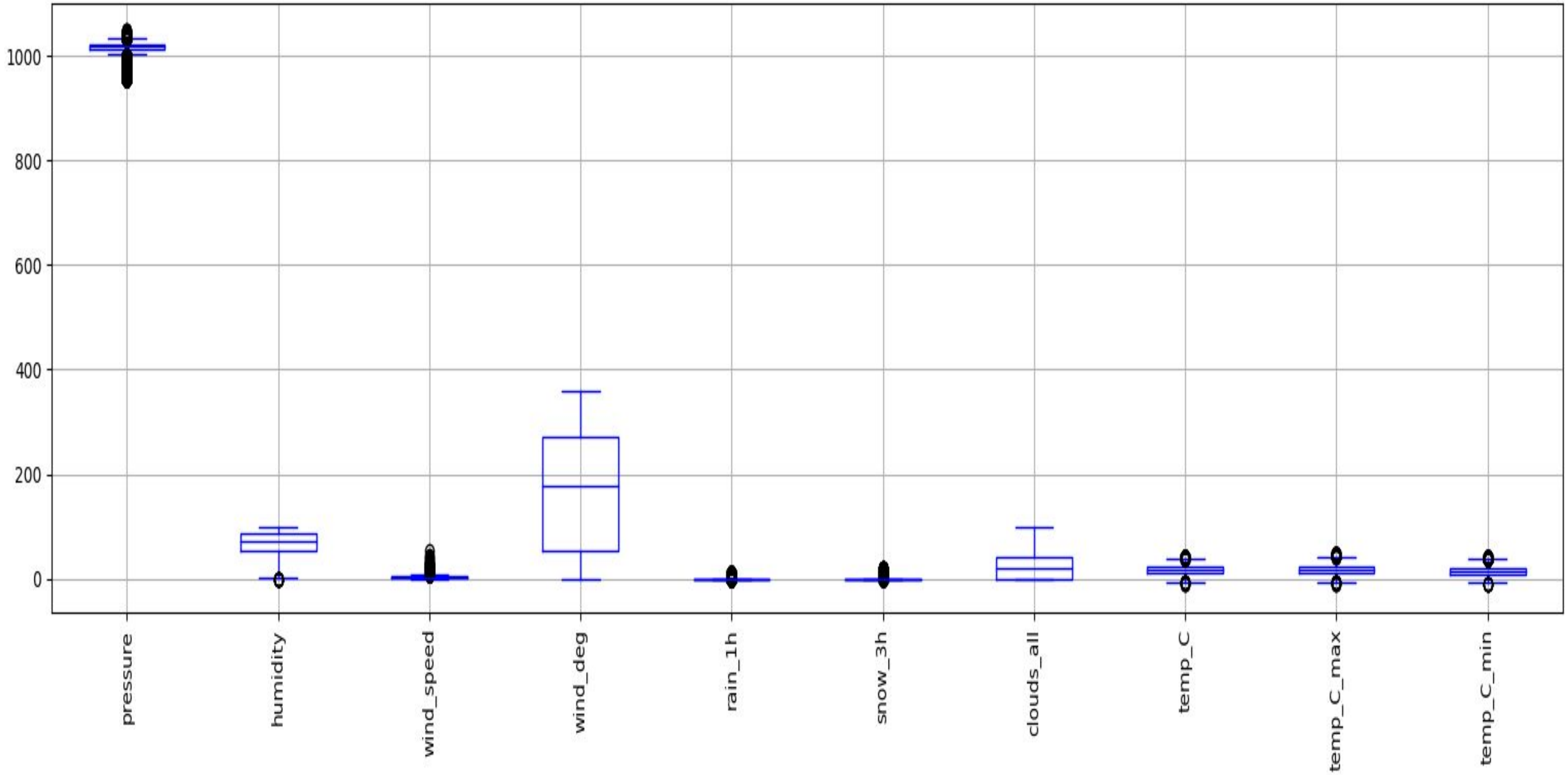
Ryerson University

Box Plots



Ryerson University

Box Plots Continued



Correlation Matrix

Attribute	**Correlation Coefficient with Target**
pressure	0.01
humidity	**-0.37**
wind_speed	**0.20**
wind_deg	-0.09
rain_1h	**0.02**
snow_3h	-0.01
clouds_all	0.01
temp_C	0.20
temp_C_max	**0.20**
temp_C_min	0.21

Correlation Matrix

Attribute	Correlation Coefficient with Target
generation biomass	**0.08**
generation fossil brown coal lignite	**0.28**
generation fossil gas	**0.55**
generation hard coal	**0.40**
generation fossil oil	**0.50**
generation hydro pumped storage consumption	-0.56
generation hydro run of river and poundage	**0.11**
generation hydro water reservoir	**0.48**
generation nuclear	**0.09**
generation other	0.10

Correlation Matrix

Attribute	Correlation Coefficient with Target
generation other renewable	**0.18**
generation solar	0.39
generation waste	0.08
generation wind onshore	0.04
total generation	**0.81**
forecast solar day ahead	**0.40**
forecast wind onshore day ahead	0.04
total load forecast	1
price day ahead	0.47
price actual	**0.43**

Selected Features for Regression

- **total generation**
- **humidity**
- **wind_speed**
- **rain_1h**
- **temp_C_max**
- **generation biomass**
- **generation fossil brown coal lignite**
- **generation fossil gas**
- **generation hard coal**
- **generation fossil oil**
- **generation hydro run of river and poundage**
- **generation hydro water reservoir**
- **generation nuclear**
- **generation other renewable**
- **forecast solar day ahead**
- **price actual**

Predictive Modelling

- Experimental design consist of 80:20 random split of data into training and testing
- Data frame for modelling consist of 16 predictors and 1 target column
- Target variable is **total load actual** measured in Megawatts (MW)
- The total number of rows in dataframe are 34468
- Therefore 27574 of 80% of records are for training and rest are for testing
- For Q1, data is sampled on hourly basis while for Q2 and 3 it is resampled on daily basis
- After resampling to daily basis the number of rows reduced to 1460

Regression Analysis - Hourly Period

Model	RMSE	R^2	Rank
Linear Regression	1748	0.85	3rd
KNN Regression	1430	0.90	2nd
Decision Tree Regression	1693	0.86	4th
Random Forest Regression	1155	0.94	1st

Regression Analysis - Daily Period

Model	**MAPE**	**RMSE**	**R^2**	**Rank**
TSO Forecast	0.008	3.09	-	**1st**
Linear Regression	2.93	1042	0.86	**2nd**
KNN Regression	3.49	1296	0.78	**4th**
Decision Tree Regression	3.81	1479	0.71	**5th**
Random Forest Regression	2.86	1058	0.85	**3rd**

Ryerson University

Classification Analysis

- Target : high_load_level
- Binary classification
- Almost no class imbalance
- Data is normalized

Ryerson University

Classification Analysis - Daily Period

Model	Accuracy	Precision	Recall	Rank
Logistic Regression Classifier	93.5%	93.5%	93.5%	2nd
KNN Classifier	90%	90%	90%	3rd
Decision Tree Classifier	100%	100%	100%	1st

Conclusions

- Creating time feature improve accuracy slighty
- There is demand shortage over 4 year
- Cross Validation can be done as a future work

References

- **David Et Al. retrieved from https://arxiv.org/abs/1906.05433**
- **Kaggle retrieved from https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather**
- **Entsoe retrieved from https://transparency.entsoe.eu/dashboard/show**
- **Esios retreived from https://www.esios.ree.es/en/market-and-prices?date=19-05-2023**
- **Openweather retrieved from https://openweathermap.org/api**

Thankyou for Listening

Ryerson University