Dialog Data Science Academy

Daily Energy Utilization Prediction

Capstone Project Report

Asoke Rathnayake
5-15-2022

Table of Contents

1. Introduction

ABC Company Private Limited is a leading retail supermarket chain in Sri Lanka. They have over 100+ retail outlets which build state of art modern supermarkets in the context of the Island wide. Last few years they have also focused to reduce their operational overhead in a different context and identified power and energy as a bigger component in that.

They need to build an AI/ML data analytical and prediction solution for their energy utilization and anomaly detection. In the first phase, they are expecting a daily energy utilization of a selected outlet which they have planned to integrate with their energy management solutions and enterprise resource planning system.

This report describes the solution which builds for their requirement and also the process and solution approach and machine learning model building.

2. Data set for solution development

ABC Company Private limited provide a data set of one of their retail outlets in Colombo which is operated from 08:30 AM to 10:30 PM. They have already implemented remote Realtime energy monitoring and collecting data in 1 min intervals.

However, for the model building and evaluation, the energy management solution has provided a data set of 12 months which is a summary of daily energy usage. And ABC Company private limited expects a daily energy usage prediction.

The data set doesn't have any null or outliers, because the energy management system provided the summary data set.

The below table provides the description or definitions of the dataset.

Field Name	Description
RefNo	Record number of the data set
Date	Date
Year	Year 2020 - 2021
Month	Month 1 to 12
Day	Day of the month 1 t o31
WeekDay	day of the week 1 to 7
KWH	Kilo watt-hours / number of units consumed in the relevant day- The unit of measure is Kwh

The below table shows sample data of the evaluation.

	RefNo	Date	Year	Month	Day	WeekDay	KWH
140	141	3/17/2021	2021	3	17	4	2155.638
239	240	6/24/2021	2021	6	24	5	1321.234
217	218	6/2/2021	2021	6	2	4	1595.281
170	171	4/16/2021	2021	4	16	6	2144.164
283	284	8/7/2021	2021	8	7	7	2109.402

3. Methodology

The Solution building had done with the Cola Python notebook. However, planning to deploy the live solution on the cloud in a python environment. API will be published via API gateway which is already available with ABC Company private limited.

a. Solution Approach

- Identify the data set and what represent
- Identify the validity of the data set
- Clean or cleanse the data set
- Identify the field/columns data distribution
- Identify the core relation of the fields with Daily energy usage
- Build the solution
    - Define the function to evaluate different models
    - Check and select the most suitable model
    - Save, test, and publish the model as API

b. Exploratory Data Analytics

The below table illustrates the field, data types, and the number of records.

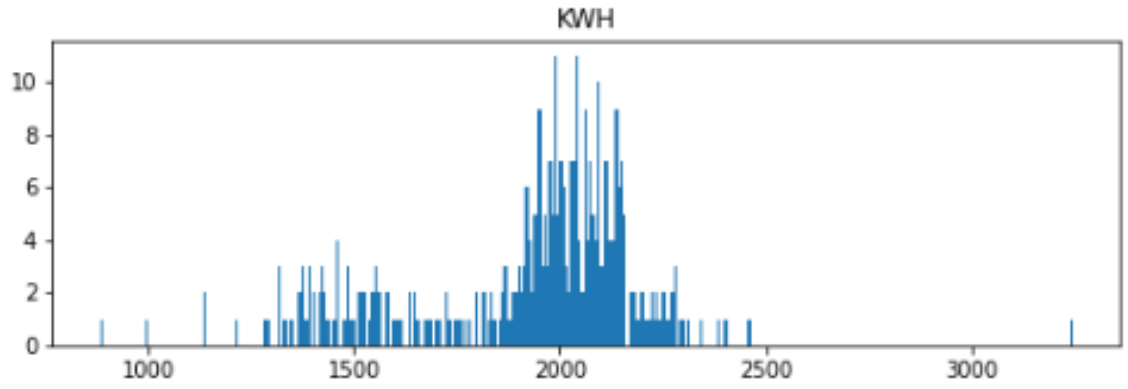
```
RangeIndex: 366 entries, 0 to 365
Data columns (total 6 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   RefNo    366 non-null    int64
 1   Year     366 non-null    int64
 2   Month    366 non-null    int64
 3   Day      366 non-null    int64
 4   WeekDay  366 non-null    int64
 5   KWH      366 non-null    float64
dtypes: float64(1), int64(5)
```

The data set didn’t have any null or missing values, which were confirmed in the notebook.

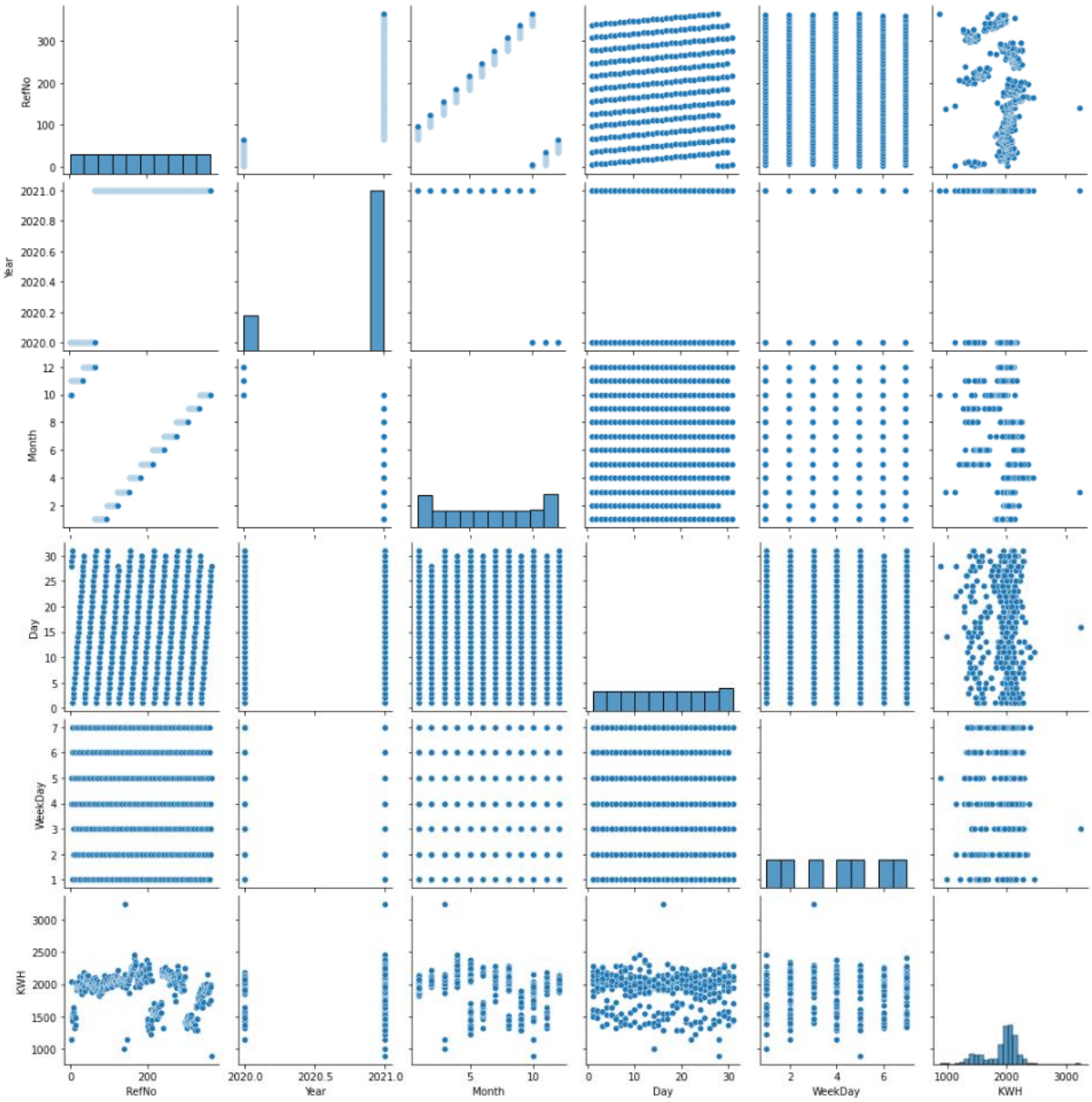
```
data.isna().any()
```

```
RefNo      False
Year       False
Month      False
Day        False
WeekDay    False
KWH        False
```

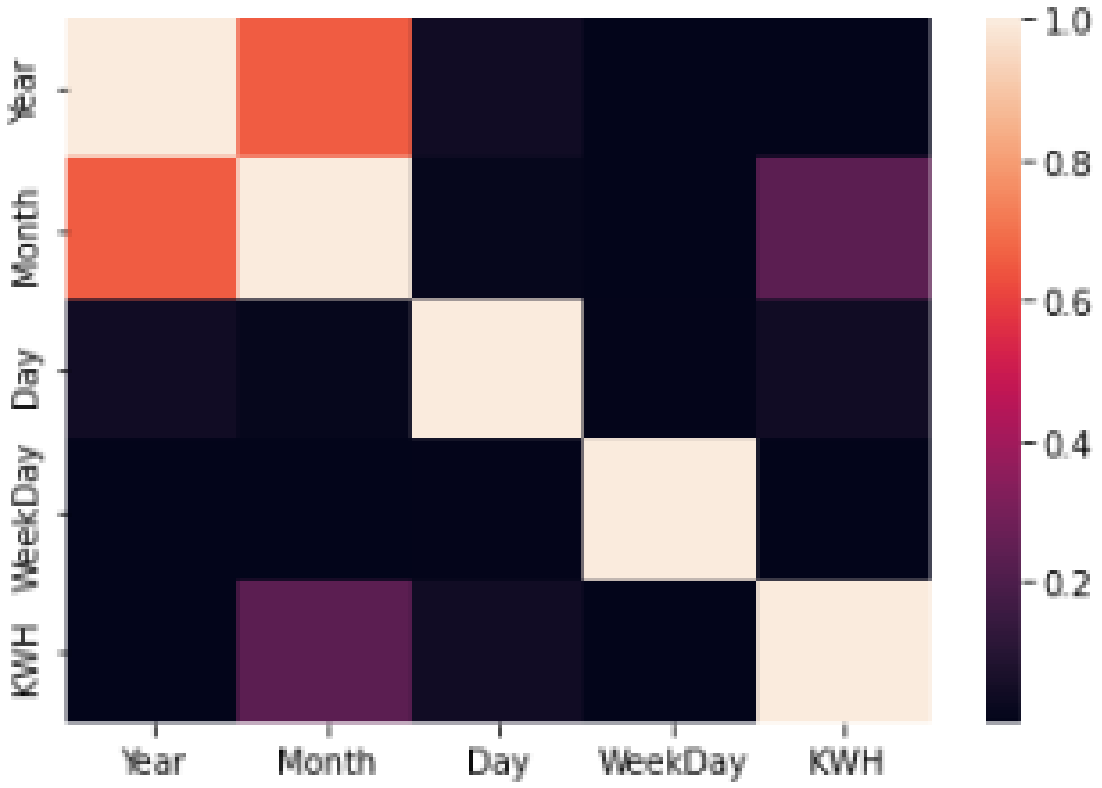
Below histogram illustrate how data get distributed specially KWH (daily energy utilization)



The Seaborn pair plot uses to check the correlation with each parameter.



To get an idea of how each parameter/feature related to each other plot a heat map



4. Feature Engineering

Feature engineering is a key task in ML model building, however, the data set provided here is having limited attributes. There is a categorical variable “WeekDay’ which consists of the day of the week. however, that attribute already comes with numeric categorical values.

5. Model Building

When model building, we have created a function for Model evaluation, a function to summarize the model evaluation result.

And also created a function for data pre-processing and cleansing. Then Publish the trained model as API, which can be directly integrated with energy management solution.

6. Conclusion

Note: Due to some errors I couldn't manage to evaluate the models via function. However, I've tried the Linear Regression model by applying test train data without a function.

Model evaluation outcome of MSE, RMSE, and R2

```
Intercept = 2328.528096856552
MSE: 93847.72730289168
RMSE: 306.3457642972915
R2: 0.05165261094078777
```