DOCUMENTATION

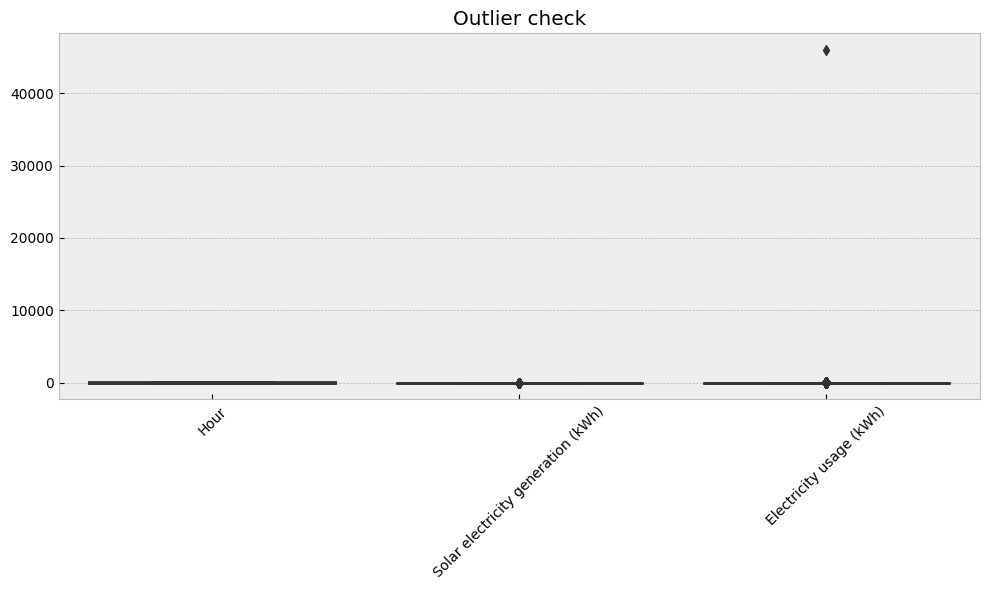
**SOLAR ENERGY AND ELECTRICAL USAGE ANALYSIS**

1. **DATA UNDERSTANDING**

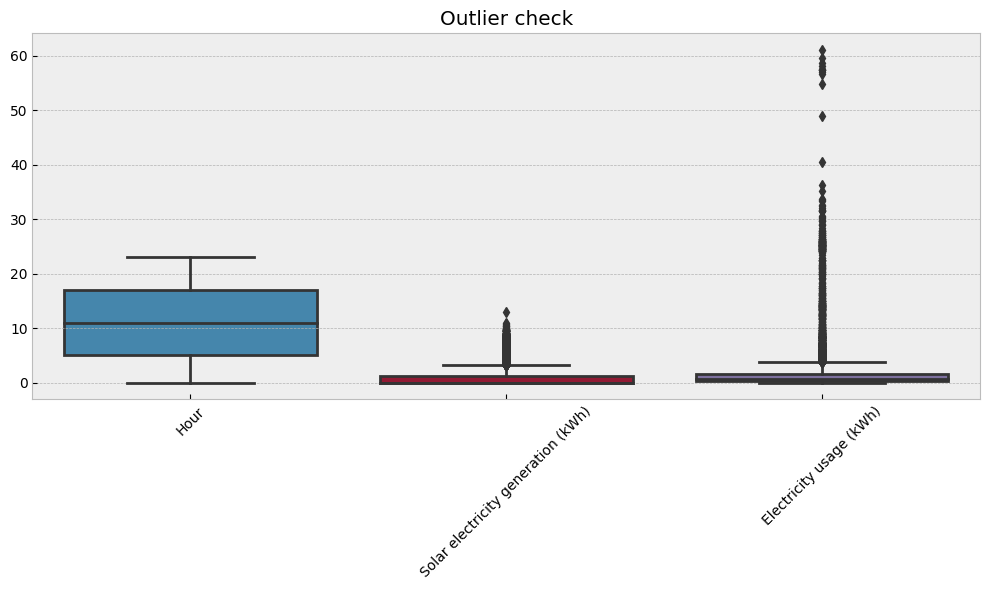
* The data comprises of historical data from the year 2020 and encompasses solar energy generation and electrical usage per hour.
* The data had 8760 rows and 3 columns.

1. **DATA CHECKS**

* The data set was checked for duplicate and null values and none were found
* Outliers were singled out in the energy usage column and these were deleted from the dataset
* Below is the initial visualization of the outliers using box plot



* This is after the outliers were dealt with :



* The outliers were dropped because:
  + In comparison with the rest of the dataset, electrical usage of 46,000 kwh was too high
  + There are no negative values for electrical usage as this is scientifically improbable in relation to electricity.

1. **Model Purpose**

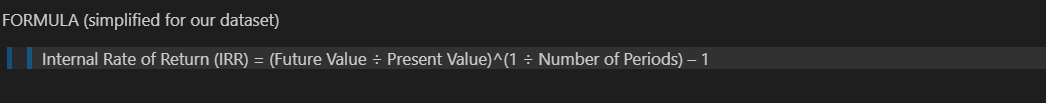
* The model used was intended to predict annual cost savings over a twenty year period.
* FB Prophet was used to achieve this.

1. **Methodology and Plan of Approach**

* The model data frame was grouped by month and the battery level columns since the savings are a derivative of battery level column multiplied by the cost of purchase of electricity.
* The logic behind this is that anytime the battery level was not zero, it signified electricity purchased forgone hence translates to a saving of that cost (foregone cost).

1. **Calculations and Feature Engineering**

* Month and Year were extracted from the date time column.
* The new column was then grouped and a summation of battery level throughout the month was generated
* From the scenarios, the new battery level now forecasted over a twenty year period, was used to calculate cost saving for both the scenarios
* Compounded cost was first calculated and this multiplied by battery level to reach the annual cost savings.
* Net Present Value was calculated from the new cost saving figures in both scenarios.
* IRR was then generated with the new columns using the formula below.

****

1. **Assumptions made**

* The discounting rate remained constant throughout the twenty year period
* No external factors save for seasonality affected the data set
* This would include purchase or sale and use of new electrical appliances, number of family members will stay the same, and factors such global warming, damages, repairs and maintenance will not affect solar energy generation nor the electrical usage.
* Apart from the outliers the rest of the dataset provided is accurate or has minimal errors