**Purpose of the Model**

The model extensively explores the data of Junior Analysts, delving deep to uncover hidden patterns within the dataset. It not only identifies these patterns but also extracts vital insights that hold significance for both Stakeholders and senior management within the organization.

Employing a systematic and methodical approach, the model meticulously guides through the process in a series of intuitive Python notebooks (. ipynb). These notebooks are crafted with clear and insightful comments, designed to be comprehensible to individuals with varying levels of expertise. Whether you're a technical expert or someone new to the field, the commentary will facilitate your understanding, making the analysis accessible and transparent.

The step-by-step breakdown of the analytical journey ensures that every stage is well-documented, making it easy to follow the reasoning behind each decision. The model's transparency and comprehensibility ultimately empower stakeholders and decision-makers to grasp the intricacies of the analysis, fostering informed discussions and strategic decisions based on the derived insights.

**Data and Checks**

1. **Setting up the Environment**

To kickstart the project, a Jupyter Notebook named "electricity\_consumption" was created. The environment setup included the following steps:

* Installation of Jupyter Notebook using the command `**pip install jupyter`**
* Launching Jupyter Notebook with **Visual Studio Code**
* Creating a new notebook and renaming it to "electricity\_consumption"
* Importing essential Python libraries for data analysis:
  + **`import pandas as pd`**
  + **`import numpy as np`**
  + **`import os`**
  + **`import matplotlib.pyplot as plt`**
  + **`import seaborn as sns`**
  + **`import plotly.express as px`**

1. **Data Importation checks**

Before delving into the analysis, it's crucial to ensure that data has been imported accurately and conforms to the anticipated format. This section outlines the steps taken to import the data and perform preliminary data checks. The data was imported and stored in an object named **electricity\_consumption\_data**.

* We parsed the **Date\_hour\_start** to **datetime** format to prevent unnecessary errors downstream.
* The first two rows were skipped since they did not have important data and were making it difficult to read the header
* The data was restricted to 8760 rows because for some reason pandas was reading the entire sheet

1. **Unifying naming conventions**

All Standardization of column naming conventions to lowercase was executed to ensure code consistency and maintainability, mitigating the potential for errors arising from mixed naming styles.

1. **Checking if the data is Complete and fit for use**

This validation process involves the following steps:

* **Statistical Measures**: Employed the `**describe()`** function to obtain rapid insights into the statistical attributes of numerical variables in the dataset.
* **Data Information**: Utilized the `**info()`** function to gain an understanding of the dataset's characteristics, encompassing data types and memory usage.
* **Missing Values**: Leveraged `**df.isnull().sum()`** to identify and quantify null values; fortunately, this dataset exhibited no missing values.
* **Outliers Handling**: Employed the boxenplot to visualize dataset outliers, followed by Z-score-based outlier adjustment and the trim method for removal. Subsequent boxplot checks confirmed successful outlier removal

**Assumptions Used**

During the calculation of the cumulative battery charge level for each hour over 2020, it is assumed that the battery stores limitless electricity

**These sections effectively explain various aspects of the data analysis**

**Checking the amount of electricity to be bought for each hour in 2020 from the electricity provider (measured in kWh and subject to a minimum of zero)**

The data is first grouped by ‘date\_hour\_start’ and aggregated by sum of electricity\_usage\_kwh

**Checking the excess solar Electricity generated**

This will be achieved by finding the difference of the solar electricity generated and the electricity used (usage).

**Checking the cumulative battery charge**

This will be achieved by summing the electricity available at the beginning of the year with net for each hour

**Finding the amount of electricity for each hour in 2020 that would have been bought from the electricity provider (measured in kWh and subject to a minimum of zero), assuming a battery had already been installed.**

To achieve this the we first pick the net negative data from the net\_electricity column (feature) and summing up their absolute values

**Finding the Savings Over 2020 from installing a battery compared to using the existing solar panels alone**

We first take the absolute value of the bought electricity and find the difference with net positive from the net\_electricity column (feature), thereafter you multiply with the cost $ 0.17 per kWh

**Building Bespoke Charts**

**-**We first build a bar graph for monthly solar generation; this is achieved by grouping the data by month with regards to solar electricity generation and aggregating by sum.

-The second bar graph is monthly solar usage; this is achieved by grouping the data by month with regards to solar electricity usage and aggregating by sum

-The third bar graph is monthly electricity purchased from the electricity provider (without battery); this is achieved by grouping the data by month with regards to solar electricity purchased (net negative) and aggregating by sum

-The last bar graph is monthly electricity purchased from the electricity provider (with battery); this is achieved by grouping the data by month with regards to solar electricity purchased (net negative) usage and aggregating by sum.

* In the last two Bar graphs there is no inherent difference since the net negative does not rely on the presence or absence of a battery

**Projection**

Using the NPV to calculate the projection over the 20 years

**Analysis for Naomi’s Data**

Calculating NPV and IRR for two different scenarios

**Exporting The Data for Further Visualization and Reporting**

The already cleaned and feature engineered data is exported to be analyzed in Business Intelligence tools (Power BI / Tableau)

**Project repository:** *https://github.com/Ernestcollins15/PULA\_assessment*