**Module 6 Assignment – Individual Project Proposal Analysis and Edits**

Sunil Raj Thota

College of Professional Studies, Northeastern University

**Executive Summary**

Studer Innotec's products are adapted to new technologies with smooth operations and with the highest safety standards. The most efficient is by far the most effective. Studer is, in my opinion, the best reliable and cost-effective inverter/charger in the marketplace currently. One of the problems that this company is facing is data gathering and performing analytics. Also, measuring, monitoring, and tracking electrical checks are lacking in their systems. To address all these issues, we need to develop a detailed presentation. This presentation will include visualizations, data insights, and recommendations.

Data collection and analytics are two issues that this organization is dealing with. Electrical checks are also not measured, monitored, or tracked in their systems. We'll need to create a full presentation to cover all of these difficulties. Visualizations, data insights, and suggestions will be included in this presentation. Studer Innotec, for example, provides global backup solutions for all industries that require a stable and uninterruptible power supply for their appliances to complete complex procedures.

**Goal and Objectives**

The main goal is to optimize the solar energy delivery and check if there is a way to reduce the genset costs. The ethical issues stem from the activities we accomplish (or don't do) with data. Data gathering, security, and utilization all necessitate ethical decisions. From the information provided by the sponsor, our team has decided to provide insights to the sponsor based on the following business questions:

* Total Power (kW) and Energy (kWh) Delivered to Installation
* Total Power (kW) and Energy (kWh) Consumed
* Power (kW) and Energy (kWh) from Solar Panels
* Power (kW) and Energy (kWh) from Genset

Identifying, accessing, and exploring the data they need, when they need it, and how they use it is essential. Most of the challenges and issues can be solved by maintaining data ethics with appropriate data stewardship by teams, he says.

**Project Scope**

Utilizing the data provided by the sponsor, Studer Innotec, and the tools at our disposal, such as Tableau and Python, we conducted an analysis using the data provided by the sponsor. Our main goal would be to use Python to parse the log files. We're still working out the details of the data we'll be using for this project. Modeling would be out of scope because the business questions our team is working on do not necessitate the development of any predictive models.

**Tools and Techniques**

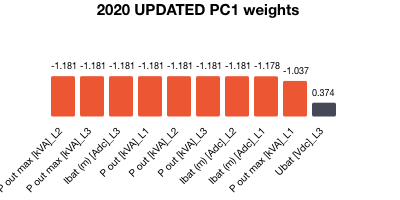
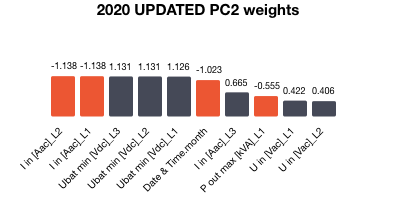
Python in Jupyter Notebook and Tableau for the sponsored project is utilized. To develop an interactive dashboard as requested by the sponsor, we will be using Tableau. The parsed data from the log files will be fed into Tableau to develop the dashboard. BigML provides a wide range of fundamental Machine Learning resources that can be combined to address complex Machine Learning problems. Regardless of whether you have any prior experience with Machine Learning, the introductory tutorials below will help you get up to speed with the BigML Dashboard.

**Data Preparation & Exploratory Analysis**

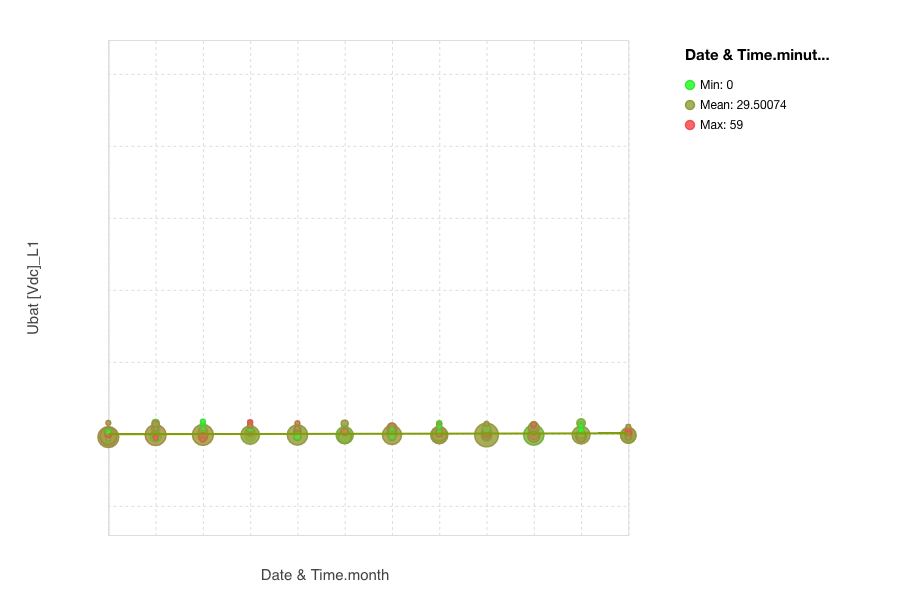
Multiple folders in the Studer Innotec dataset contain various log files, Xtender, and Messages data in CSV format. 1866 Log files, 18 Message files, 16 RCC files, 1 VT file, and 48 XTENDER files are included. We've chosen to work with the log files provided, but because there are so many, we've decided to start with only a handful, to begin with, and add more as the project goes.

The log files' data will be processed, and any null values will be erased. The data will be correctly prepared to ensure that all columns have the correct data type. The timestamp column, for example, will be divided into Date and Time columns, with the Date column, further divided into Year, Month, and Day. By splitting the data in this way, we'll be able to go more specific with our answers to the business questions. To forecast these factors, I will use the BigML Tool. I'll be employing supervised learning techniques such as classification, regression, decision trees, random forests, and time-series forecasting. Because of the BigML forecasting models, which feature a clear user interface as well as modeling and analysis components that are simple to comprehend. These can be downloaded and used on any edge local computer to provide simple, offline predictions, and they can be deployed instantaneously as part of global, real-time production systems.

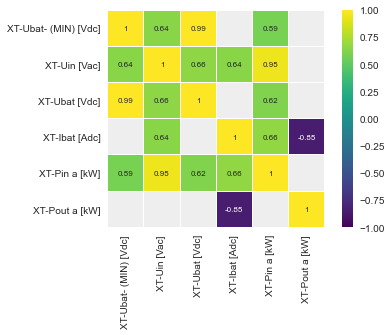
**Data Visualizations**

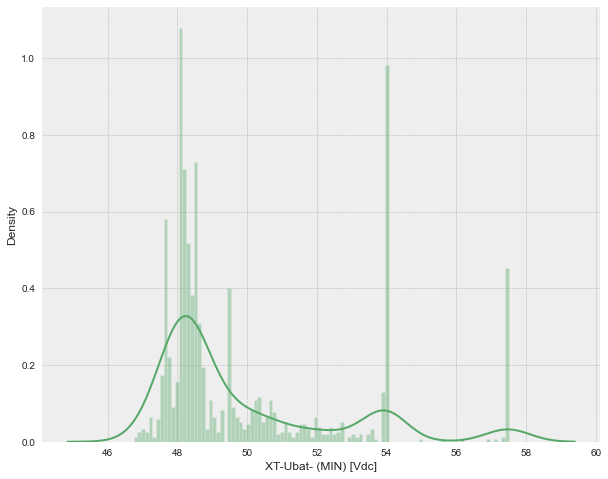
The above visualizations are used to depict the principal components of the data. As it simplifies the complexity in high-dimensional data while retaining the patterns and trends. It has been performing on the Year 2020 related log files. The next visualization is between the Ubat [Vdc]\_L1 and the Date & Time in terms of months. It is a scatter plot with a Pearson coefficient of 0.08 and a spearman value of 0.13439.



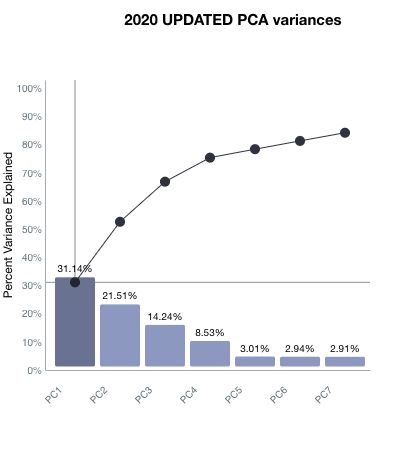
The above signifies the correlation plot between the columns of the dataset. I also made a heatmap to show how each feature connects to the others. A lighter hue indicates a better correlation between the two features, whereas a darker color indicates a weaker correlation.



The above plot showcases the density distribution of the XT-Ubat-(MIN) [ Vdc] Column which tells about the Minimum Batt Voltage [V] from Xtender #1. We can see that most of the Min Batt Voltage are in 47 to 49.



The below graph is related to the principal components percent variance in %’s



**Failed Approaches**

As the dataset consists of 1850+ CSV log files it has been a tedious and time-consuming process to work with Excel, Python, and R. As all these have failed for me, BigML had helped a lot to load the data and work on the EDA part. In this, we can use ML-related models also and it is way simpler than the other tools.

**Findings and Recommendations**

PC 1 with 34.1% is leading in the principal components table. The solar energy power in and power out of the solar panels have a low positive connection (0.07). Solar panel efficiency is not high, at least according to the present dataset. A solar panel's ability to consume more solar energy does not imply that it will produce more electricity. P in KW and Pout KW are both present, but there is no granularity in the overall visual, i.e., which systems contributed the most, i.e., L1/L2/L3. In the visual presented, there was no indication of which systems were investigated, and an excellent visual practice is to always keep a label for the plots to identify the parameters.

**Conclusion**

The sponsor’s business question for this project entailed understanding how Studer Innotec would optimize the delivery of solar energy. Studer Innotec aims to ensure renewable energy sources like solar adoption in all areas, including remote areas. For the company to achieve its objective, it must understand the power and energy delivered by specific products like solar panels and genset generators. The techniques used to tackle the sponsors’ business questions will be Tableau and Python applications. To foresee the repercussions of power and energy difficulties, I'd like to use machine learning approaches. To study and present the data, I'll use the Visualization Dashboards. Consequently. Look for data about the weather that is readily available and download it. Create a vivid account of the consequences of carbon dioxide emissions and analyze them.

**References**

Impulse, S. (2021). *Studer Innotec Sa - a member of the World Alliance*. SolarImpulse Foundation. Retrieved October 17, 2021, from https://solarimpulse.com/companies/studer-innotec-sa.

Studer Innotec. (2021). *NEXT3*. STUDER LIVE | next3. Retrieved October 18, 2021, from https://studerlive.ch/next3/.