**Module 10 Assignment – Individual Project Proposal: Draft Submission**

Sunil Raj Thota

College of Professional Studies, Northeastern University

Dr. Matthew Goodwin

**Executive Summary**

Studer Innotec's products are designed to adapt to new technologies while maintaining smooth operations and the greatest levels of safety. By far the most effective is the most efficient. Studer is now the most reliable and cost-effective inverter/charger on the market. 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. They'll need to create a detailed presentation to handle all of these difficulties. Visualizations, data insights, and recommendations will be included in this presentation (Impulse, S., 2021).

This organization is concerned with two issues: data collection and analytics. Electrical checks aren't measured, tracked, or monitored in their systems, either. To address all of these issues, they'll need to produce a comprehensive presentation. This presentation will feature visualizations, data insights, and recommendations. For example, Studer Innotec provides global backup solutions for all industries that demand a stable and uninterruptible power supply for their equipment to finish complex tasks (Studer Innotec, 2021).

**Objectives and Scope of the Project:**

The goal of this project is to develop a model and dashboard that can forecast battery power distributions. The major goal is to improve the delivery of solar energy and see whether there is a method to cut genset expenses. Ethical concerns arise from the things they do (or don't do) with data. Data collection, security, and use all require ethical judgments. Other parameters would be included in the suggested model for evaluating the value of the battery analysis. This could be due to a lack of understanding of how to use data efficiently or how to disperse resources to regions where they are truly needed.

* 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. Utilizing the data provided by the sponsor, Studer Innotec, and the tools at their disposal, such as Tableau and Python, conducted an analysis using the data provided by the sponsor. Our main goal would be to use Python to parse the log files. Worked out the details of the data they'll be using for this project.

**Design and Data Collection Methods**

Internal resources can be used to acquire data, or external resources can be used to purchase the data needed. Using logs as a starting point for data collecting might be an excellent place to start. Another efficient method is to conduct user surveys of existing customers or to include a brief survey for new consumers who register (Müller, Michael & Bründlinger, Roland & Arz, Ortwin & Miller, Werner & Schulz, Joachim & Lauss, Georg, 2014).

An examination of research publications could potentially reveal customer preferences. Raw data in the form of logs will constitute the majority of the data collected. This is related to the Studer Innotec dataset, which has completely different data. Another aspect of the initiative is to examine past data and use predictive analytics approaches to forecast the value of comparable ways in the future. The data will be cleansed and prepared so that it only contains important information for modeling. The organization can evaluate how to effectively utilize the Genset and Battery Generation based on this information (Evolution, 2021).

**Modeling Methodologies**

Employ both descriptive and predictive analytics for this project. Before starting any modeling or research, it's important to figure out what elements influence the value of a battery's manufacturing and generation. This will necessitate a thorough examination of the data as well as the market itself. We will gain greater insight into the market by researching competition marketing strategies, consumer engagement approaches, and investments. Another aspect of the endeavor entails reviewing previous data and applying predictive analytics techniques to forecast the value of a similar piece. Customers can be retained using predictive modeling techniques by offering them individualized analyses and insights.

**Tools and Techniques**

For the facilitated project, Python in Jupyter Notebook and Tableau are used. Tableau will be used to create an interactive dashboard as desired by the sponsor. Tableau will be used to create the dashboard using the parsed data from the log files. BigML offers a variety of basic Machine Learning resources that can be integrated to solve larger Machine Learning tasks. The basic courses below will help you get up to speed with the BigML Dashboard, regardless of whether you have any prior expertise with Machine Learning.

**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, they'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, they'll be able to go more specific with their answers to the business questions (Khatibi, Arash & Bendfeld, Jörg & Bermpohl, Wolfgang & Krauter, Stefan, 2017). To forecast these factors, will use the BigML Tool. They will 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.

**Preliminary Studies or Pilot Tests:**

Daytime power consumption is lower than evening power consumption, according to the hourly trend of energy consumption and output. The recommendation here is to optimize power generation so that more energy is produced during peak hours. In 2015 and 2020, the amount of power generated by diesel generators per hour is similar. Pout Max (KVA) L1 has a significantly greater value than Pout Max (KVA) L2 and Pout Max (KVA) L3 after cleaning the null value.

The idea is to look at these developments throughout five years (2015–2020). Increasing the efficiency of the equipment by optimizing power generation according to trends. Because the other phases aren't being used as much, Pout Max (KVA) L1 should be considered. All of the data analysis and visualization capabilities are crucial in understanding how Studer Innotec can achieve its solar power optimization aim. Tableau is one of the data visualizations tools that can be used to provide insights into the best approach to ensure solar optimization. Furthermore, the usage of photovoltaic panels has allowed for cost savings on electricity purchases from the national grid, resulting in a not-too-high payback period.

**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, BigML had helped a lot to load the data and work on the EDA part. In this, they 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. This one-of-a-kind solution allowed the farm's trout to be fed without having to move big bales on trolleys every day or rely on other forms of external power that would have made trolley handling more difficult.

**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 methods used to tackle the sponsors’ business questions will be Tableau and Python applications. To study and present the data, let’s use the Visualization Dashboards.

**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.

Müller, Michael & Bründlinger, Roland & Arz, Ortwin & Miller, Werner & Schulz, Joachim & Lauss, Georg (2014). *PV-off-grid Hybrid Systems and MPPT Charge Controllers, a State-of-the-Art Analyses*. Energy Procedia. Retrieved December 01, 2021, from http://dx.doi.org/10.1016/j.egypro.2014.10.133.

Khatibi, Arash & Bendfeld, Jörg & Bermpohl, Wolfgang & Krauter, Stefan. (2017). *Testing and Analysis of Battery Charge Controllers for Off-Grid PV Systems*.

Kroposki, B.; Pink, C.; Lynch, J.; John, V.; Daniel, S.M.; Benedict, E.; Vihinen, I.; *"Development of a High-Speed Static Switch for Distributed Energy and Microgrid Applications,"* Power Conversion Conference - Nagoya, 2007. PCC '07, Conference Record, pp.1418-1423, 2-5 April 2007.

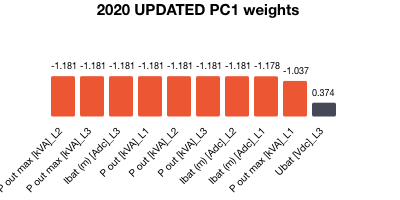
Studer Innotec. (2021). *NEXT3*. STUDER LIVE | next3. Retrieved October 18, 2021, from https://studerlive.ch/next3/.

Evolution, M. (2021, January 29). *How to Use Predictive Analytics in Data-Driven Marketing*. Modern Marketing Measurement & Optimization. https://www.marketingevolution.com/knowledge-center/the-role-of-predictive-analytics-in-data-driven-marketing.

**Appendix A: Data Visualization Results**

**Figure 1**

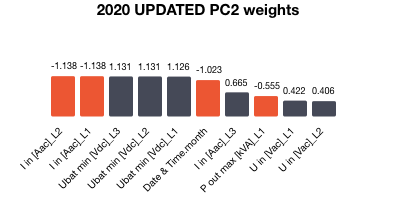
*2020 Updated PC1 Weights*



*Note.* 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.

**Figure 2**

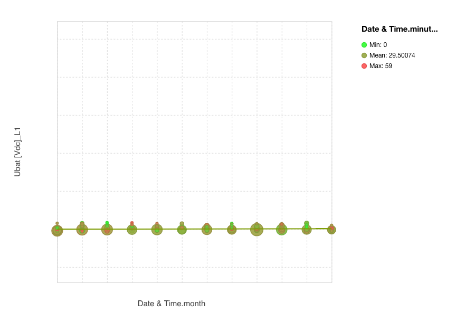
*2020 Updated PC2 Weights*



*Note***.** It has been performing on the Year 2020 related log files.

**Figure 3**

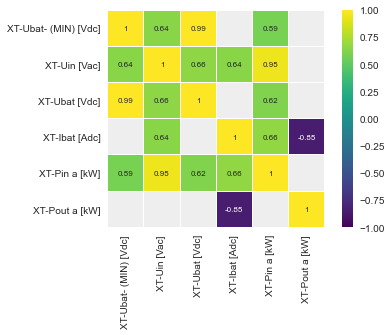
*Ubat [Vdc]\_L1 vs Date/ Time*



*Note***.** This 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.

**Figure 4**

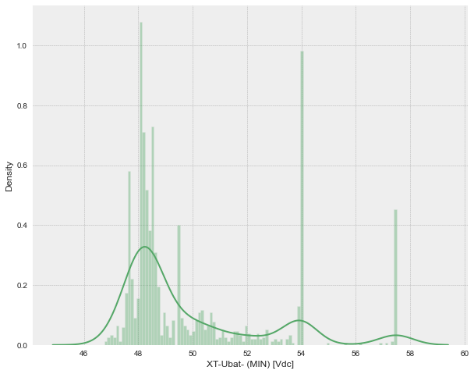
*Correlation Plot*



*Note***.** A lighter hue indicates a better correlation between the two features, whereas a darker color indicates a weaker correlation.

**Figure 5**

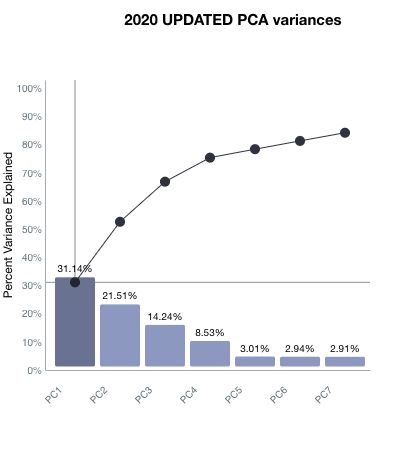
*Density Distribution XT-Ubat-(MIN) [ Vdc]*



*Note***.** 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.

**Figure 6**

*2020 Updated PCA Variances*

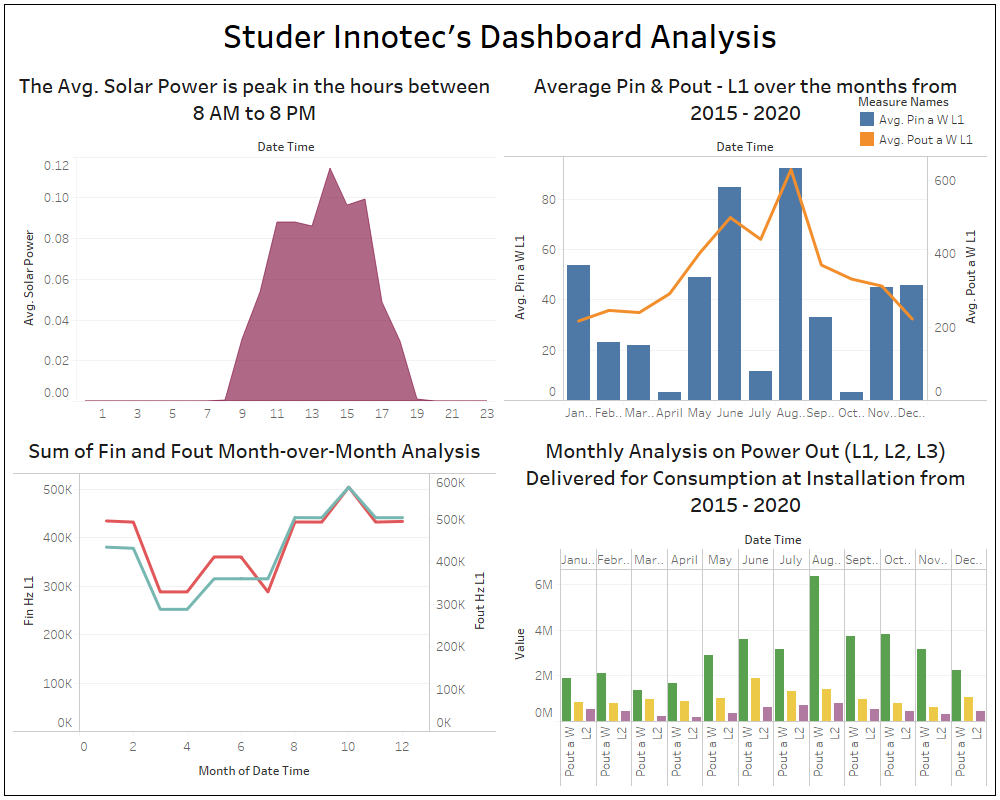


*Note***.** The above graph is related to the principal components percent variance in %’s

**Appendix B: Tableau Dashboard**

**Figure 1**

*Studer Innotec’s Dashboard Analysis*



*Note.* Average Solar Power is the total solar energy/ Power In from the Solar Panels. This second diagram depicts the amount of solar energy that enters and leaves the solar panels. We can observe there is a peak in October where the Fin and Fout are the same. It's the monthly analysis on Power Out delivered for consumptions at Installations from 2015 to 2020.

**Appendix C: Literature Review**

Kroposki, B.; Pink, C.; Lynch, J.; John, V.; Daniel, S.M.; Benedict, E.; Vihinen, I.; *"Development of a High-Speed Static Switch for Distributed Energy and Microgrid Applications,"* Power Conversion Conference - Nagoya, 2007. PCC '07, Conference Record, pp.1418-1423, 2-5 April 2007.

Kroposki (2007) found that “In lesser, small mini-grids with centralized generating, these systems are often the simplest to deploy. The issues get more difficult when energy levels rise and the network grows bigger and much more distributed regionally. A decrease in the accessibility of off-the-shelf products and parts heightened effective communication between network components increasing concern about maintaining voltage control over the mini-grid” " (p. 5). More R&D of control components and systems for massive, dispersed mini-grids is required. New data analysis techniques frequently lead to algorithms that can aid in the prediction of energy utilization and off-grid systems. Advanced analytics is producing a lot of excitement amongst scholars and scientists, and it's being used in a multitude of sectors at a quick pace. They want to investigate new predictive analytics innovations and how they impact decision-making in energy firms like Studer Innotec.