**Discussion 1: Introduction to Capstone Research**

Post a summary of one of your literature reviews and how it informs your capstone project.

Off-grid power sources are among the most effective strategies to address the globe's increasing energy demands in developing nations and remote communities while also reducing harmful fossil fuel emissions. Energy firms are attempting to improve their heterogeneous catalysts in order to lower infrastructure costs as well as yearly maintenance costs in order to compete effectively in the industry.

People remain to rely on fossil fuels for energy, which is warming the earth and creating drastic global warming. All across the summers of each year, media sources recorded multiple incidents of excessive rainfall, floods, and scorching in the United States. According to the US Energy Information Administration, fossil fuels accounted for nearly 80% of primary energy demand and 93 percent of carbon dioxide emissions in 2018. (Source: EIA, 2020.)

The world energy revolution is centered on decreasing energy-related greenhouse gas emissions. The globe must transition away from fossil fuels, which are at the root of climate change, and toward sustainable energy sources such as solar and wind. Solar and wind energy investments can help to enhance quality of the air. In Asia and Europe, air pollution continues to be a major public health concern. Uncontrolled, ineffective, toxic fossil fuel transportation, industry, and power stations are the primary causes of air pollution. Switching to sustainable renewables will suitable in the long haul while also reducing pollution and preserving the environment in places around the world.

**References:**

[1] Huneke, F., Henkel, J., Gonzalez, J.A, Erdmann, G. (2012). Optimization of hybrid off-grid energy systems by linear programming. Energy, Sustainability and Society 2.1 (2012): 1-19.

[2] Global Investors Move into Renewable Infrastructure. ieefa.org/wp-content/uploads/2021/07/Global-Investors-Move-Into-Renewable-Infrastructure\_July-2021.pdf.

**Discussion 2: Exploring the Data and Searching for Literature**

Now that you have completed some research in the domain area of your sponsor, how does that information inform your interpretation of the business question or sponsor issue? What are intermediate questions that will provide a pathway to a solution that your sponsor is looking for?

Transportation, industry, household, and commercial use of fossil fuels are the most significant. There could be a lack of knowledge and understanding regarding the need of solar energy optimization. The strategy, as well as criteria that do not always fit the regional socio - economic and environmental circumstances, are the primary reason for poor renewables use. Many places require off-road automobiles to travel about, that makes bringing goods and commodities in operationally difficult. The project's goal is to find strategies to improve solar electricity distribution while lowering fuel usage. Data inputs will be a critical component of the project's final product. Solar energy models were the topic of simulation model, which included cash benefits, legislation, and temperature fluctuations. The sites where Studer Innotec chooses to deploy will determine how much fuel is saved. The sponsor's approach seems unlikely to coincide with the company's plan to focus on regions where fuel consumption is minimal.

If the dashboards has to be detailed or can accept a higher tier strategy for reporting out of questions that need answering. This leads to a need for a complete grasp of the dashboard's stakeholder’s identification. Other groups, such as management and operation, may demand various degrees of data or sorting features, for example. Collaboration with the sponsor will be improved by developing intermediary questions. The inquiries will center on the issue's core reasons, remedies, and alternative responses. What adds to energy and fuel waste, for example, are among the intermediary issues. How could electricity use be tracked in order to avoid loss? Also, we can develop few recommendation systems or any machine learning models to predict the patterns and find some critical insights that might help the Sponsor.

**References:**

[1] Best, S. (2011). Remote access: Expanding renewable energy provision in rural Argentina through public-private partnerships. London: International Institute for Environment and Development

[2] P R, Krithika & Palit, Debajit. (2013). Participatory Business Models for Off-Grid Electrification. 10.1007/978-1-4471-4673-5\_8

**Discussion 3: Draft a plan on how you envision tackling your project proposal (the business/research question you are proposing for the Sponsor to implement).**

This project's purpose is to assist Studer Innotec in increasing solar power output from their energy conversion devices. Data processing, data filtering, data cleansing, EDA, data modeling, feature selection, and evaluation techniques will be employed to answer the sponsors questions. First, machine-learning techniques are increasingly being used to renewable energy, with artificial intelligence techniques and hybrid models accounting for the bulk of predictions in solar and wind energy. Second, for machine-learning models in renewable-energy projections, the decomposition method is used more frequently than other data pre-processing techniques. Finally, extreme-learning machines and support-vector machines are two of the most common machine-learning models that use metaheuristics to pick parameters. By performing these techniques Studer Innotec can analyze, control, and maintain solar power conversion technologies, as well as redistribute future generating power output, by predicting the future solar power energy.

I will also intend to aggregate all time-series files in order to visualize and predict the target variable across several geographies and platforms. The business aspect that interests me is how Studer Innotec obtains data and where they preserve it for inside researchers to access. The focus is on building analytics solutions and keeping data securely location increases the speed with which data can be pulled for planning and strategic findings. Also, I will be focusing on the problem that we are trying to solve in the project. I will list out the deliverables and keep note of the benefits of the proposal. I will define a dedicated timeframe for the project and how effectively meet the deadlines. Also, it is necessary for all of us to find and solve the risks and potential issues involved in tackling this project with the teams.

**References:**

[1] Lai, J.-P., Chang, Y.-M., Chen, C.-H., & Pai, P.-F. (2020). A Survey of Machine Learning Models in Renewable Energy Predictions. *Applied Sciences*, *10*(17), 5975. doi:10.3390/app10175975

**Discussion 4: Sharing Proposal Research - Post a review of ONE PEER REVIEWED article on the theme of your project proposal.**

The S2G-Home is a domestic smart grid facility that will be used to test and evaluate a variety of decentralized load management technologies. The control systems we're working on have been placed in twenty homes on a low-voltage residential grid in Switzerland's south. Prior to implementation in pilot households, the S2G-Home allowed for the validation of the energy management algorithm and the created hardware solution in a controlled environment. A PV generator, two battery-to-grid systems, a synthetic load, and an electric vehicle supply device were all included in the test facility.

The algorithm's first real testing revealed qualitatively plausible behavior, including the ability to adapt to unanticipated voltage fluctuations. The testing also confirmed the requirement to ignore voltage drops caused by algorithm decisions in order to ensure a higher quality inference of the network's future state and achieve smooth functioning. The algorithm can be deployed in the pilot neighborhood when the control system has been validated in a test hardware environment.

MPC is predicted to become the control method of choice for power electronic systems with several coupled variables and tight operating limits that operate at low pulse counts. The reader will learn how to increase the converter's power capability, lower current distortions, reduce filter size, achieve very fast transient responses, and ensure reliable operation within safe operating area constraints using Model Predictive Control of High-Power Converters and Industrial Drives.

**Peer Reviewed Article:**

[1] Rivola, Davide and Medici, Vasco and Giusti, Alessandro and Rudel, Roman (2014) A Residential Smart Grid Facility for Testing and Evaluation of Decentralized Load Management Strategies. In: EUPVSEC 2014, 29th European PV Solar Energy Conference, 20-25.09.2014, Amsterdam, The Netherlands. https://repository.supsi.ch/5549/

**Discussion 4: Annotated Bibliography Part 2**

**Post the summary of your fourth or fifth article and how those peer-reviewed papers inform/contribute to your project proposal.**

Renewable energy generation methods have been employed to meet increased electricity demand while also lowering greenhouse gas emissions. Among the several renewables, solar is the most viable alternative. When compared to other energy sources, however, the photovoltaic system only converts 30–40% of direct solar radiation into electricity. Muller (2014) found "Due to battery capacity and costs, renewable energy is crucial in all off-grid operations. All possible efforts to reduce power demands and improve part performance must be done to establish a stable and cost-effective network". Due to the non-linear features of solar PV and environmental conditions, the performance of solar PV is substantially reduced (p. 4).

Filtering circuits can help to prevent this. Energy electrical devices such as DC-DC converters and inverters receive the resulting signal. When lighted, a photovoltaic (PV) network is a combination of solid-state semiconductor technology that generates energy. A photovoltaic cell is one of the components of a solar panel. Several solar cells are connected in series and parallel to form a solar component. PV arrayed in series to provide the best electrical output and in parallel to achieve the best output.A considerable study has been undertaken for a long time to verify the effectiveness of PV systems and to investigate the various challenges involved with the optimal exploitation of solar PV systems in the process of extracting the maximum output from them. The MPPT controller's tool generates oscillations as a result of the closed-loop system monitoring of solar light. Solar forecasting can be done with global energy as an input parameter or with its components: directed radiation, diffused radiation, and reflected radiation.

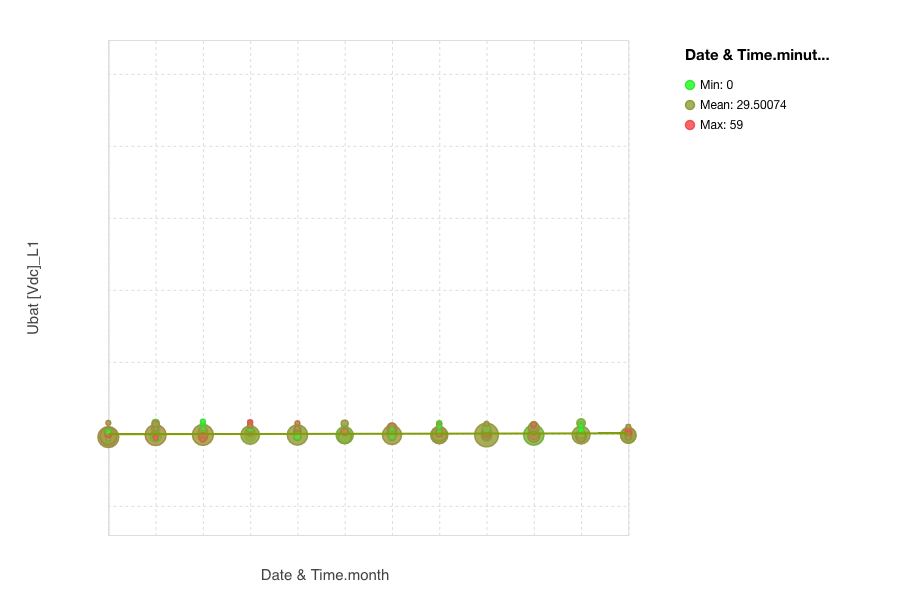
**Peer Reviewed Article:**

[1] 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. http://dx.doi.org/10.1016/j.egypro.2014.10.133

**Discussion 5: Data Sets and Graphs**

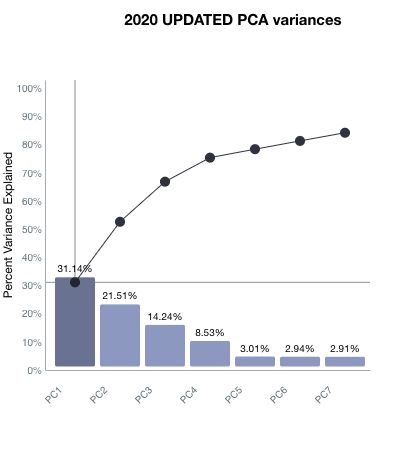
Post a screenshot of at least one data set you will use on your Individual Project Proposal and the screenshot of a graph that will be part of your analysis. Give a quick summary description of both.

I had incorporated a 2020-year data from Studer Innotec's log files dataset for the individual project proposal and group project. The files must first be converted from CSV to Excel, with the comma delimiter acting as a column separator. After the data has been integrated, it can be ingested into Python or Tableau to generate business insights. As part of my initial Exploratory Data Analysis, I accessed the combined and cleaned data in Excel before making the switch. I have used BigML to do some analysis and work on the plots.



The above plot signifies the relationship between Ubat [Vdc]\_L1 and the months in the year 2020. We can observe that the mean is 29.5. If the dashboards has to be detailed or can accept a higher tier strategy for reporting out of questions that need answering. This leads to a need for a complete grasp of the dashboard's stakeholder’s identification.

The dataset we used for the capstone project, which was provided by the sponsor Studer Innotec, is shown below. It's a log file for operations, with data points collected every second. We loaded log data for one year, 2020, for this project. There are approximately 0.5 million rows of logging data supplied, which is enough to cover two years and aids in the discovery of seasonal patterns.



The above plot is a representation of the Principal Component Analysis percent variance in %’s. The PC1 component has the most 31.4%. PCA helps us interpret the data, but it will not always find the important patterns. It simplifies the complexity in high-dimensional data while retaining trends and patterns. It does this by transforming the data into fewer dimensions, which act as summaries of features

**References:**

[1] Lai, J.-P., Chang, Y.-M., Chen, C.-H., & Pai, P.-F. (2020). A Survey of Machine Learning Models in Renewable Energy Predictions. *Applied Sciences*, *10*(17), 5975. doi:10.3390/app10175975

**Discussion 8: Data Ethics:**

Read the articles under Supporting Material and reflect on what ethics means in the realm of big data and predictive analytics and how it applies to your project proposal. When possible, please include references to articles on Ethics in the area of your proposal.

We must ensure that our energy infrastructure is sustainable, goal-oriented, and operational as part of our broader energy and climate strategy. On the supply side, increased efficiency can be achieved by better managing fluctuating renewable energy generation options, improving energy grid infrastructure maintenance, and even improving demand modeling and, as a result, infrastructure investments. On the demand side, more efficient energy consumption patterns can be achieved by real-time demand-response pricing and grid load balancing (Hatzakis, 2019).

Voltage and frequency fluctuations induced by the intermittent nature of renewable energy production owing to changing weather conditions, as well as harmonic (wavelength) distortions introduced by electrical devices used in renewable energy generating, can cause quality difficulties. There appears to be an underlying suspicion of energy companies and governments in terms of their attitudes and pledges to policies that will benefit society as a whole if they are implemented.

Data Ethics in our Project that we follow and abide by our Customer information and identities should be kept secret. Private information that has been shared should be kept private. Customers should be able to see everything. Big Data should not be allowed to obstruct human decision-making. Unfair biases should not be institutionalized by big data.

Smart meters can be used to reveal detailed information about people's private lives within the home, raising serious questions at a technical and policy level about in-home surveillance and how to address consumers' privacy interests. Expected consumer benefits, such as energy savings, are predicated on the collection and analysis of granular information on household energy usage.

**Reference:**

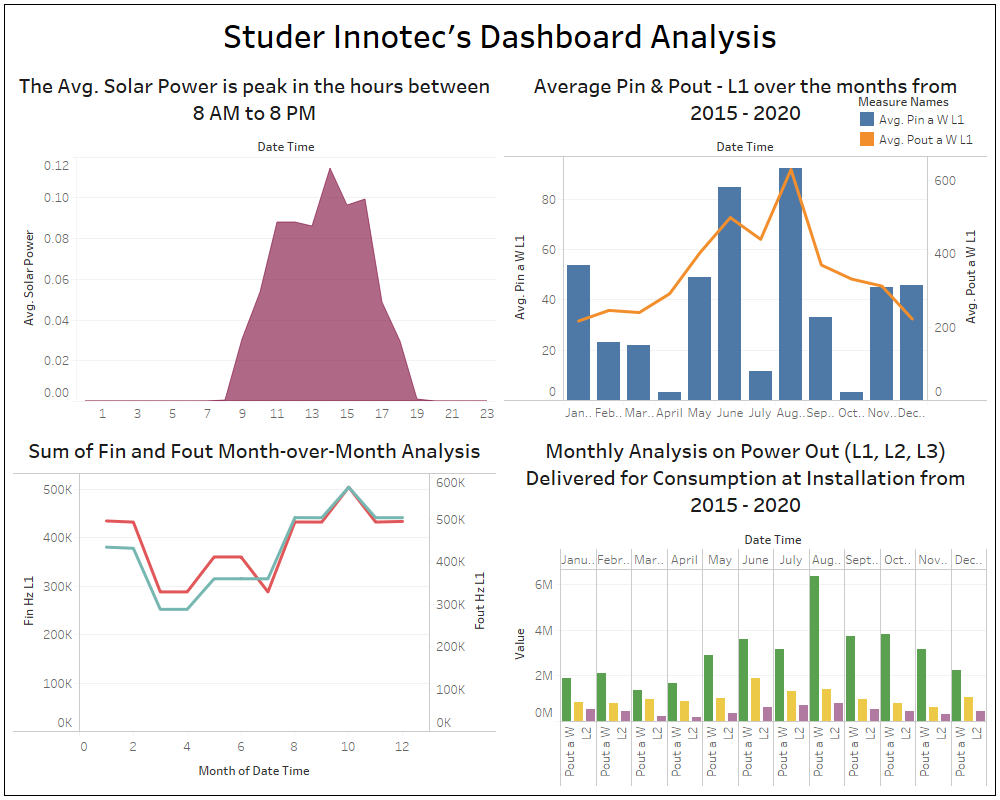
[1] Hatzakis, T., Rodrigues, R., & David, W. (2019). Smart Grids and Ethics. *ORBIT Journal, 2*(2). https://doi.org/10.29297/orbit.v2i2.108

**Discussion 9: Capstone Check-in:**

Post the screenshot of the dashboard or model you (individually) are working on for the CAPSTONE and give a quick summary of it.

Making sure we're using the proper chart/graph type for your data is the first step towards effective data visualization. Here are some nice overviews of our options and why some are better in particular scenarios.

This summary of data graphing options presents the issue in terms of the underlying statistical topic. Are you, for example, comparing numbers? Do you want to show your composition? Are you looking for the latest fashion trends? Are you looking for a distributor? The type of graph you use to depict and explain the results should be informed by what you're attempting to learn from your data.



Here is the screenshot of the Dashboard that I was working on for the Capstone.

* The first graph depicts that the Average Solar Power is at a peak in the hours between 8 AM to 8 PM. It is the total solar energy/ Power In from the Solar Panels.
* The second graph shows the relationship between Power In and Power Out Max in kW for one of the transponders. This second diagram depicts the amount of solar energy that enters and leaves the solar panels. It has little effect on the amount of solar energy that comes in and out of the panels.
* The 3rd plot shows us the minimal difference between the fin and Fout over the 12 months from 2015-2020. We can observe there is a peak at October where the Fin and Fout are the same.
* The fourth chart tells the comparison between the Pout among L1, L2, L3. It’s the monthly analysis on Power Out delivered for consumptions at Installations from 2015 to 2020. The least L3 is observed in April month.

**Reference:**

[1] Knaflic, Cole Nussbaumer. (August 14, 2020)."Exploratory vs Explanatory Analysis." *Storytelling with Data*, http://www.storytellingwithdata.com/blog/2014/04/exploratory-vs-explanatory-analysis

**Discussion 10: Data Governance:**

Discuss the considerations regarding data governance on your individual project proposal. Is that an issue? If so, how will you explain the issue and propose a solution or strategy? Provide a summary so your peers can comment and help you to craft a solid section on your proposal about data governance. You should have read the book cited in Modules 2 and 6. Make sure to add sources in APA format to back up your ideas. Below is a short white paper and a blog that might help you as well:

Data governance is a process that combines data collection, policies, standards, and measurements to guarantee that information is used effectively and efficiently to help an organization achieve its objectives. The sponsor's data collection mechanism is, in my opinion, highly mature; although I have no idea how effective the procedure is, the offered variables are detailed enough to characterize all features of a solar PV system. The data records, on the other hand, were collected in minutes, which could result in data storage concerns and the possibility of data loss, as well as increasing the standards of physical analytical equipment. The enormous amount of data size necessitates not only a high-performance computer to play and alter the data, but also increases the processing time for changing, lowering the efficiency of using this data.

The data will be acquired and specified with a framework to put it into before the experiment is actually executed due to the nature of the experiment. The solar panel, angle/tilt, azimuth angle (the angle of the sun), cleanliness, and wind of the solar panel throughout different times of the day are the metrics being measured and recorded in the experiment to maximize energy output. In this respect, the data is predetermined and strictly governed by the experiment design.

Organizational construction, data security, data development, data quality, metadata management, and data value are typically included in the scope of a data governance project. Of course, Studer's execution of the core path of data governance is the same, regardless of the diverse goals of each company's data governance. My particular data governance aim for the project right now is to make it easier for the business to immediately comprehend and interpret the data, as well as to assess the efficiency of solar power generation, lighting, duration, angle, and other concerns. My individual project proposal's purpose is to capture and comprehend the optimal environmental parameters for producing the best solar PV panel output, which should reduce fuel use as well. Due to the enhanced efficiency of the solar panels, the maximum output would also minimize the use of diesel (Genset).

**Reference:**

[1] Loyens, J. (2020). Agile Data Governance: Why modern data challenges require a new approach to governance. Retrieved on November 28, 2021, from https://tdwi.org/whitepapers/2021/02/diq-all-dataworld-agile-data-governance.aspx

[2] NTT Data Research (Jul 2020). The Big Pivot: From Data Islands to Data Insights. *NTT Data*. Retrieved on Nov 23, 2021, from https://in.nttdata.com/en/insights/big-pivot/from-data-islands-to-data-insights