Aluko_Week 2: Data Science Practicum - Project Proposal

1. **Name, Contact info (e.g., email/phone):** Olumide Aluko (olumidelk@gmail.com/ 281-857-7607).

2. **Title of the project:** Predicting the Price of Electricity with Machine Learning

3. **High-level description of the project: what question or problem are you addressing?**
   Real-time electricity pricing models can potentially lead to economic and environmental rewards compared to expected flat rates. They can provide end-users with the prospect of reducing their electricity expenditures by responding to pricing that differs with different times of the day. Still, recent findings have shown that the lack of knowledge among users about how to act in response to time-varying prices and the lack of effective building automation systems are two significant barriers to fully employing the potential benefits of real-time pricing tariffs. The outcome of this project will help solve the problem as mentioned earlier by proposing an optimal and automatic residential energy consumption scheduling framework that tries to achieve the desired trade-off between minimizing the electricity payment and lessening the waiting time for the operation of each appliance in the household in the presence of a real-time pricing tariff combined with inclining block rates. The proposed project design requires minimum user effort and is centered on simple linear programming computations.

4. **What type of data science task is it?**
   - classification using supervised learning
   - prediction using machine learning
   - data visualization

5. **Data: Brief description of data. How big do you expect the data will be? Is the amount of your data too big or too small? If you're web-scraping or collecting data, how long do you expect to collect the data?** This dataset contains four years of Spain's electrical consumption, generation, pricing, and weather data. Consumption and generation data were retrieved from ENTSOE, a public portal for Transmission Service Operator (TSO) data. Settlement prices were obtained from the Spanish TSO Red Electric. I plan to download the dataset from Kaggle, and its sizes are 6,126 KB and 19, 453 KB.

6. **How will you analyze the data? What machine learning methods do you plan to use, and/or what business intelligence aspect do you plan on incorporating?** I plan to analyze the dataset with Machine Learning and using Jupiter Notebook Python 3.

7. **Describe any anticipated difficulties and problems. Discuss how you may overcome the problems:** I anticipated difficulties from updates, required libraries, error codes, Hyperparameter optimization, cleaning, etc. And I intend to research journals, stack flow, and any other material source that would guide the successful run of the dataset.

8. **Suggest a timeline for the project. This should be a weekly breakdown of what you plan on doing each week:**

   - Week 2 – Project proposal research & submission
   - Week 2 – Data collection
   - Week 3 – Data cleaning
   - Week 4 - Data exploration
   - Week 5 - EDA
   - Week 6 - Build models
   - Week 7 - Hyperparameter optimization
   - Result visualization & conclusion