Energy consumption prediction in households

Business Understanding
- What problem are you trying to solve, or what question are you trying to answer?

The problem that I am trying to solve is to predict the energy consumption in households

- What industry/realm/domain does this apply to?

Sustainable energy, Residential sectors, Commercial sectors

- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)

The motivation behind solving this problem is to help homeowners, and energy companies to optimize energy usage and save money. Also, this can promote sustainable energy usage by identifying opportunities to reduce energy consumption. And finally, it will help to prevent carbon emissions and reduce environment impact.

Data Understanding
- What data will you collect?

I will be using is SmartMeter dataset and Global Energy Forecasting dataset.

- Is there a plan for how to get the data (API request, direct download, etc.)?

Direct download

- What are the features you'll be using in your model?

Time, DayOfTheWeek, Temperature, Weather condition, HomeSpecs, Occupancy, EnergyPrice

Data Preparation
- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?

Data cleaning and normalization, Feature Engineering, Data splitting and Feature selection.

- What are some of the cleaning/pre-processing challenges for this data?

Missing Values, Outliers.

Modeling
- What modeling techniques are most appropriate for your problem?

Linear regression, Time-series Models, clustering algorithms.

- What is your target variable? (Remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)

Energy Consumption

- Is this a regression or classification problem?

Regression problem

Evaluation
- What metrics will you use to determine success (MAE, RMSE, etc.)?

MAE, RMSE and R2

Tools/Methodologies
- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?

Linear regression, Time-series Models, clustering algorithms.