Written proposal of modeling approach [Checkpoint]
Describe your planned modeling approach, based on the exploratory data analysis from the last two weeks (< 1 page, bullet points).

Data set:
The dataset contains hourly weather data specific to each wind farm contracted by Hydro-Québec, as well as aggregate hourly power production from all wind sources.

Weather data includes temperature, relative humidity, wind speed, and wind direction.

SVD suggests the dataset dimension could be reduced from 158 to 20-30 dimensions, if desired.

The aim is to use weather data and known power generation data to predict power generation one day in advance.

Choice of model:
- Time Series Dependent: for different time windows, possible inclusion of simulated weather forecast
    - SARIMA - the data is not stationary, so it needs to be differentiated (probably trend stationary)
    - SARIMAx
    - Neural Networks (Mentor Recommended) - Long Short Term Memory (LSTM), Recurrent Neural Network (RNN), adjust num layers, num nodes per layer, optimizers, Sequential vs. N-BEATS model (interpretability)?
- Non-Time Series Dependent (focusing on weather variables)
    - kNN on daily totals
    - MLR- informed by background knowledge of wind turbines, accounting for polynomial features and seasonal variations
    - Random Forest - Handles large amounts of features well, is not inherently time sensitive but can be made time sensitive by adding temporal factors like dates or seasons as features.
    - Baseline model- total mean power production per day.