# Plan for Final Capstone Project

**PART 1: EDA AND DATA PREPARATION**

**Reading in Data, with some initial data preparation**

**Exploratory Data Analysis**

* Basic plots to get an idea about the dataset
* Make some assertions based on what you see
* Begin looking for patterns
* Daily­­­­
* Weekly
* Monthly
* Quarterly
* Yearly

**More Data Preparation**

* EDA should tell us more about how we should prepare data
* Start feature-engineering (i.e weekend, months, quarter etc)
* Drop columns that are not useful
* Using DarkSKY API, get temperature date and append to dataframe

**Statistical Testing such as normality, Dickey-Fuller**

* Need to check for stationarity and normality as this is a time series
* Then skewness, kurtosis can be seen

**PART 2: NAÏVE MODELLING**

**Using Naïve Modelling to create some base models**

* What metric to use i.e. RMSE and why
* Train/test split
* Walk-Forward Validation (cumulative)
* Daily Persistence
* Weekly Persistence
* Weekly OYA

**PART 3: AUTOREGRESSIVE, MOVING AVERAGE MODELS**

**Underlying assumptions**

* Removal of trends, seasonality
* Stationarity through differencing

**Autocorrelation and Partial Autocorrelation**

* Explanations and reasons for choosing models

**SARIMAX**

* How it works
* Scale data to speed up
* Train/test split
* Grid search
* Minimize AIC and RMSE
* Produce meaningful diagrams of results

**PART 4: LSTM**

**How it works**

* Reshaping and how to actually utilize the model
* Need to choose a sequence i.e. how many hours are used to predict next hour
* Layers
* Scaling and train/test split