1.		(a) The features that I will extract from the date and timestamp are the *hour, month,* and *day of the week.* I chose the *hour* because, at different hours in the day, the energy usage can be very different. For instance, at nighttime, the energy usage will be much lower because the lights will all be off and people are asleep, but in the middle of the day there will be more energy usage because people are awake and using their appliances and increasing energy usage. The *month* and *day of the week* have been chosen for similar reasons. If the month is in wintertime, then the energy usage could be higher because people are using the heat more. The day of the week can also have an effect if people are going to work or school during the week, then less energy at the house will be used while people are out.
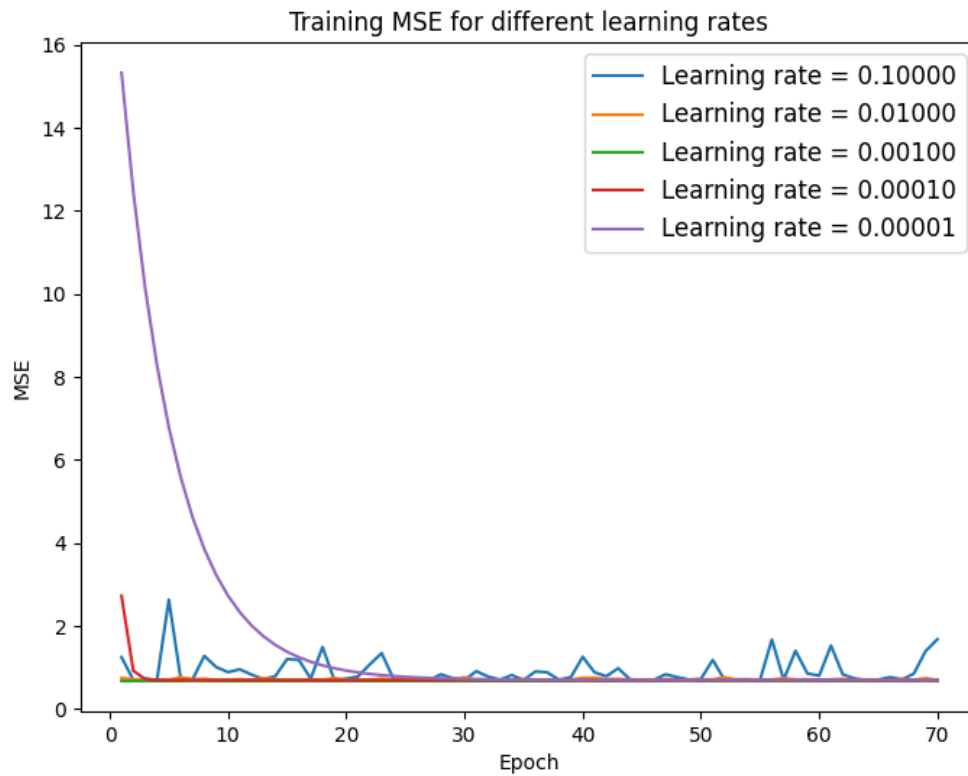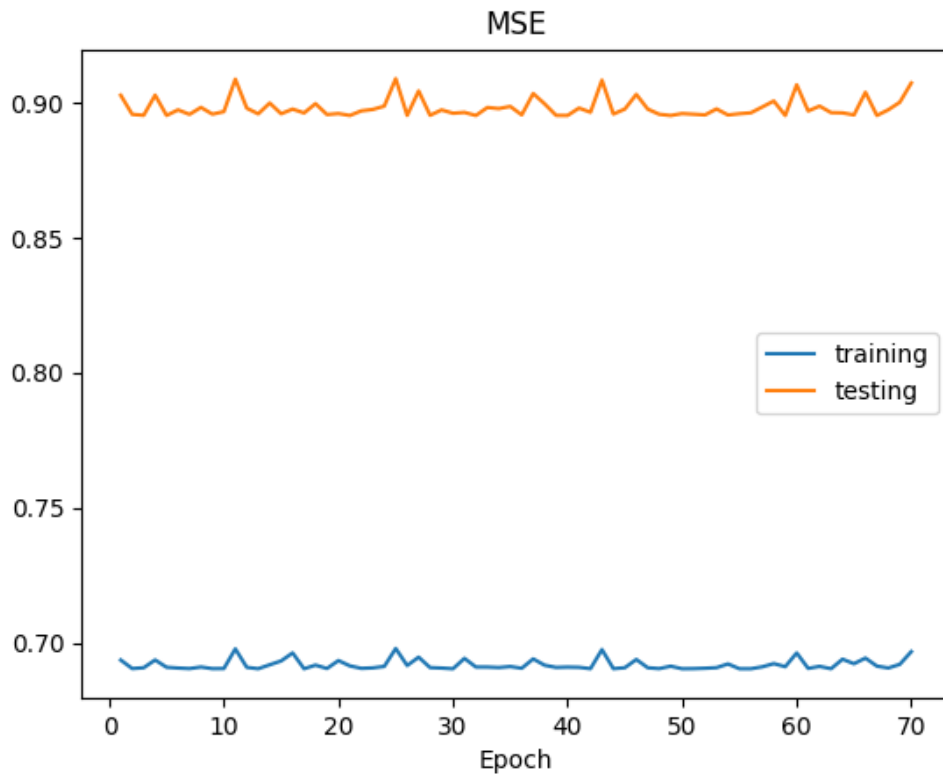
(d)

Correlation Heatmap
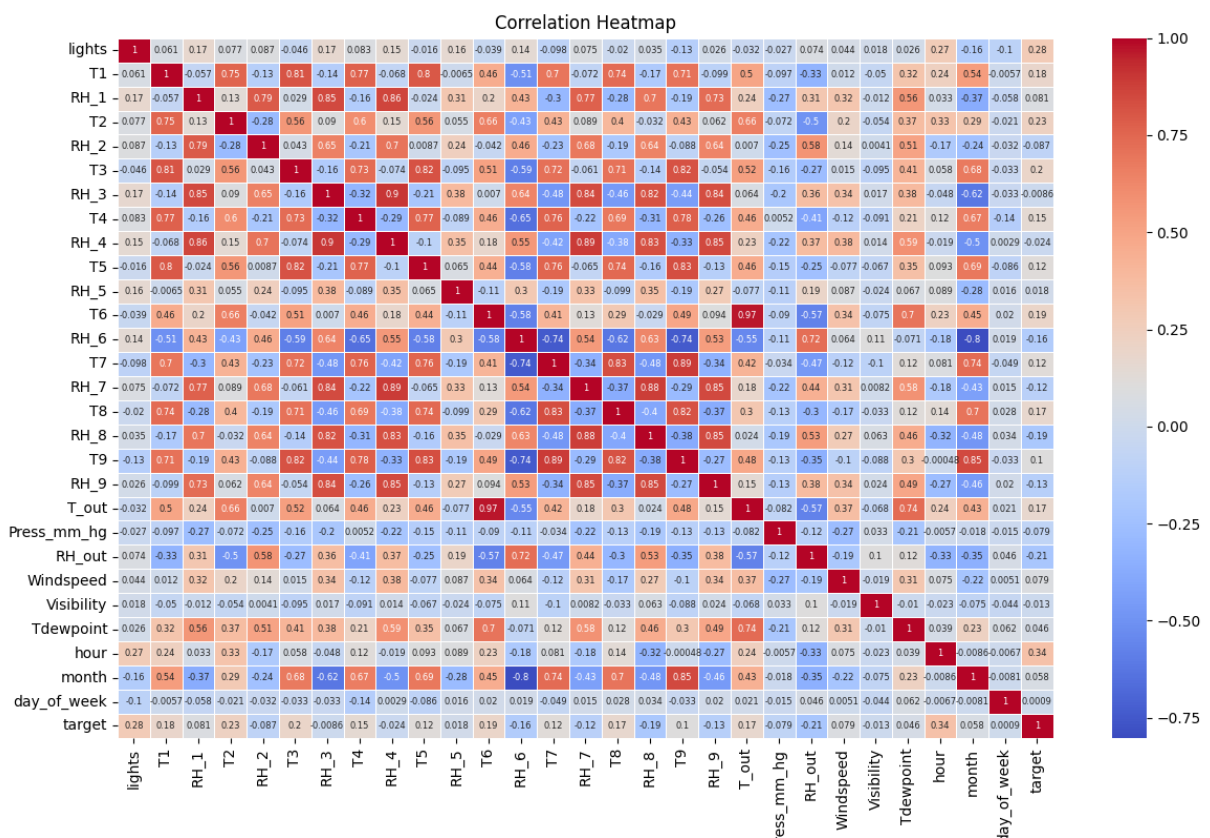
(e) My plan is to remove any features that are highly correlated with each other, which implies that they provide the similar/same explanatory power to predicting the target feature. For instance, I will drop RH_2 because it has a high correlation of 0.75 with RH_1. I will also drop T1 because it has a high correlation of 0.81 with T2. I will also drop any features that explain too little of the target feature, such as day_of_week that only has a correlation of 0.0009. Having a correlation below 0.01 is too little. I will continue doing this method, and doing so should help get rid of unnecessary features and reduce the variance and bias. After dropping features, I am left with [lights, T2, RH_2, T3, RH_4, T5, RH_6, RH_7, T8, RH_8, T9, T_out, Press_mm_hg, RH_out, Windspeed, Visibility, Tdewpoint, hour, month]

(g) Other preprocessing methods such as min-max scaling should help eliminaive the negative effect of outliers and create a scale that all the features are on.

3.   (c) Based on the plot, the optimal learning rate seems to be 0.001, because the MSE converges the fastest as the epoch increases.
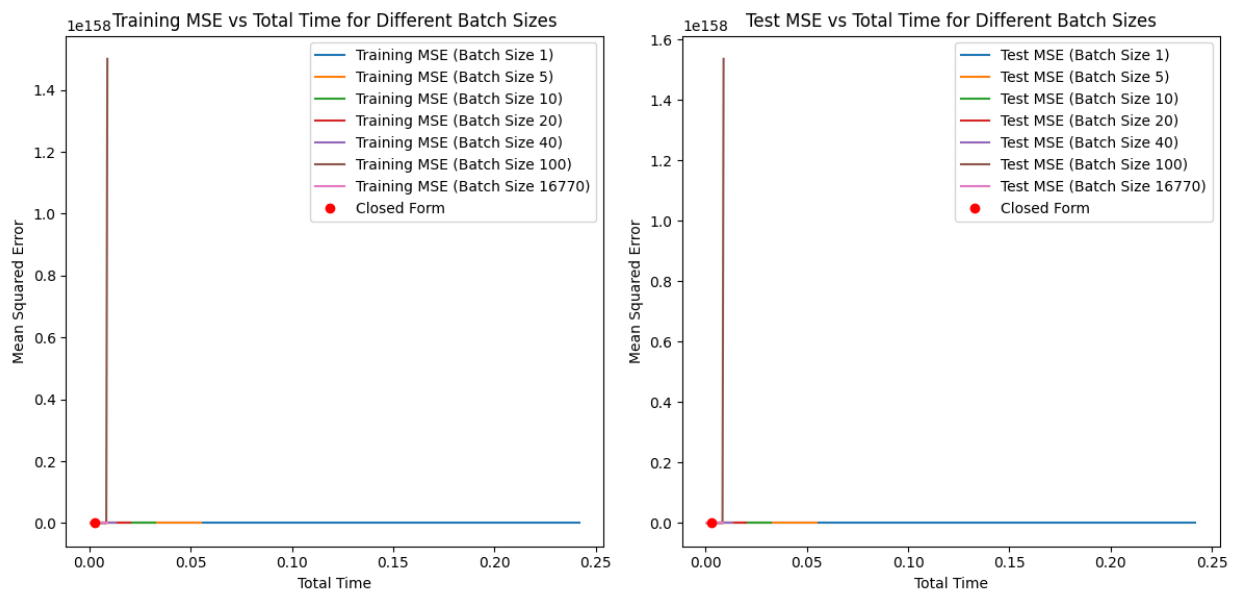
Training MSE for different learning rates

Legend:
- Learning rate = 0.10000
- Learning rate = 0.01000
- Learning rate = 0.00100
- Learning rate = 0.00010
- Learning rate = 0.00001

MSE (y-axis), Epoch (x-axis)

(d)

MSE

4.

(a)

(b) Based on the plot in (a), the closed form method outperforms the SGD in both time and MSE. The trade-offs between different batch sizes for SGD is that the larger the batch size, the longer it takes to reach its minimum MSE. Overall using the closed form method is better than SGD.