Average temperature in NYC on specific day:

R2 is very low with high SE value.

A screen shot of a graph

Description automatically generated

Yearly average temperature in NYC

R2 is still low

SE is high

The yearly average temperature and on one day are both low R2 values.

The daily average temperature is the most noisy graph as the values varies from -10 to 10.

The graph shows the global warming but with a high SE there is a change that the graph are erroneous.

A graph with blue dots and a red line

Description automatically generated

Yearly average temperature nationaly

R2 is at 72%

SE reduced

With a more cities and data, the SE is much reduced meaning that the graph would be a better representation.

Also R2 value is high at 0.72 . This helps demonstrate the global warming with a reduce margin of error.

For cities in the same region, the results would be more accurate as the temperature variation will be less. I would assume that R2 would be higher.

A graph with blue dots and red line

Description automatically generated

Yearly average temperature national with 5 year moving average

R2 is at 0.92,

SE is at 0.04

With a high R2. our model has a high level of confidence on global warming. The 5-year moving average reduced the noise.

A graph with blue dots and red line

Description automatically generated

A graph with blue dots and red line

Description automatically generatedA graph with blue dots and red lines

Description automatically generatedA graph with blue dots and a red line

Description automatically generated

All models show a upward trend.

The R2 increases with polynomial degrees increase. The model 2 with polynomial 20 fits the data the best.

A graph with red and blue dots

Description automatically generatedA graph with numbers and a red line

Description automatically generated

A graph with a red line and blue dots

Description automatically generated

R2 or RMSE is better with a degree 1. Degree 20 has a large RMSE meaning a very low accuracy between the model and the actual data.

The best model would be with 1 degree as the RMSE is very low and data a located around the line.

The difference with the previous part D.2. is that the model 20 are too complex and fit the training data too well. That is why it do not perform well on predicting data.