1. What were the initial indications from your exploration of the data that suggested that AI might add value in addressing the problem of filling in missing values in the data? Select all that apply. **1 / 1 point**

- [x] There appear to be correlations between many of the individual pollutant measurements.

  Correct
  That's right, correlations between variables, different pollutants in this case, are something an AI algorithm can learn from!

- [x] There appear to be temporal patterns in the data, like daily or weekly patterns that repeat.

  Correct
  That's right, temporal patterns are something an AI algorithm can learn from!

- [x] There appear to be spatial patterns in the data, namely that individual station measurements are consistently higher or lower depending on location.

  Correct
  That's right, spatial patterns are something an AI algorithm can learn from!

2. What are some possible approaches to replace the missing values in the data? Select all that apply. **1 / 1 point**

- [x] Use an algorithm, like a neural network, to estimate the missing value based on information from other sensor measurements as well as things like location and time of day.

  Correct
  Great job! A neural network can learn from patterns and correlations in the data to make estimates of missing values.

- [x] Copy the current measurement from the closest sensor station that is online.

  Correct
  Well done! This is a very simple method, which may be effective in some cases or serve as a baseline.

- [x] Copy the last available measurement from the sensor station that is currently offline.

  Correct
  Yes! This is a very simple method, which may be effective in some cases or serve as a baseline.

3. When it comes to designing a solution for a problem where you think AI might add value, what is a good general approach? **1 / 1 point**

- (•) Start with a simple method to establish a baseline. Then try more complex algorithms and compare them with your baseline results.
- ( ) Do some research on what AI model would best address the problem you're working on. Start with this model, and then try modifying various parameters to see if you can get an improved result.

  Correct
  Great job! Whenever possible, it's a good idea to establish a baseline. It's possible you'll find a simple solution performs well enough and no more complex approach is needed.

4. Why can a more complex method, a neural network in this case, for estimating the missing values outperform the simplest methods like copying the last recorded value or using the nearest neighbor method? Select all that apply. **1 / 1 point**

- [x] The more complex method, a neural network in this case, can learn from patterns in the data (correlations, temporal and spatial patterns).

  Correct
  That's right!

- [ ] More complex models always perform better than simple techniques.
- [x] The more complex method, a neural network in this case, can capture nonlinear relationships between features in the dataset.

  Correct
  That's right, in this case the patterns in the spatial and temporal data

5. What are some of the possible inherent challenges in accurately estimating pollution levels in between the sensors? Select all that apply. **1 / 1 point**

- [x] Pollution levels between the sensors will depend on many factors, for example, which way the wind is blowing, making a uniform distance weighting scheme limited in its practicality.

  Correct
  That's right, we didn't talk about this in the lab directly, but in general, many different factors may be at play beyond those accounted for in your model.

- [x] You don't actually have any "ground truth" measurements for pollution levels at locations where there are no sensors so any model you adopt can only be a rough estimate.

  Correct
  That's right, in the lab you used the sensor stations themselves as proxies for ground truth in estimating the error of your model but that only applies to the handful of locations where sensor stations are actually positioned within the city.

- [ ] There are not enough sensors to make a meaningful estimate of pollution in between the sensors.

7. What metric do you use in the labs to assess the performance of various models (Design phase lab videos)? **1 / 1 point**

- ( ) K Nearest Neighbors (KNN).
- (•) Mean absolute error (MAE).
- ( ) A neural network.
- ( ) Standard deviation.
- ( ) Mean squared error (MSE).

Correct
Yes! You are looking for a model that has the lowest MAE.

8. What can you say about the interpolated data between the sensor stations when using one nearest neighbor vs. three nearest neighbors? Select all that apply. **1 / 1 point**

- [x] When using three nearest neighbors, the interpolation on the map looks smoother.

  Correct
  Good job!

- [x] When using three nearest neighbors, the MAE is lower, thus the predictions can be expected to be better on average.

  Correct
  That's right!

- [ ] Distance weighting with one nearest neighbor produces a lower MAE value than distance weighting with three nearest neighbor
- [ ] The predictions are equally good when using one or three nearest neighbors.

9. What are some of the properties of artificial neural networks? Select all that apply. **1 / 1 point**

- [x] A neural network is a particular kind of machine learning model.

  Correct
  Woohoo!

- [x] An artificial neural network is made up of layers of so-called artificial neurons.

  Correct
  Yes! Many artificial neurons connected create an artificial neural network.

- [x] An artificial neural network is a computation machine that can take in a collection of inputs, run a computation, and generate an output.

  Correct
  That's right!

- [x] Artificial neural networks can generally model more complex functions of data than simple linear models.

  Correct
  Yay!

10. Which of the sentences best describes the inverse distance weighting scheme? 1 / 1 point

- You weight the measurements based on the square of the distance to the center of the coordinate system.
- You weight the measurements based on the distance from the point of interest, such that the measurements that are closer have a lower weight than the measurements further away.
- You weight the measurements based on the distance from the point of interest, such that the measurements that are closer have a higher weight than the measurements further away.

Correct
That's right!