Solutions for Data Analysis Process Quiz

Question 1 - Question Step

Given the above data on variables that potentially influence the number of bikes rented each hour, what questions would be relevant to ask?

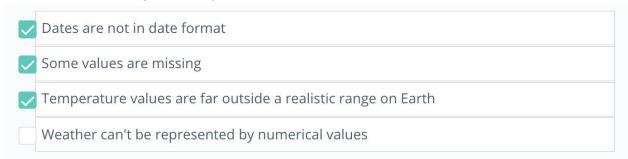
/	Which attributes are most important in predicting the number of bikes rented?
	What type of bike is rented most frequently?
/	For which day of the week should the bikesharing company run promotions if the goal i
	Should a given station stock more bikes in order to maximize profits?

Rationale:

Questions are relevant if we have the data to answer them.

Question 2 - Wrangle Step

What potential problems do you see with this Kaggle bike sharing dataset that would need to be fixed before continuing with analysis?



Rationale:

Incorrect data types, missing data, and inaccurate data are all problems that we'd need to fix before analyzing this data.

Question 3 - Explore Step

Based on these scatterplots, which of these three features seems most helpful in predicting count?

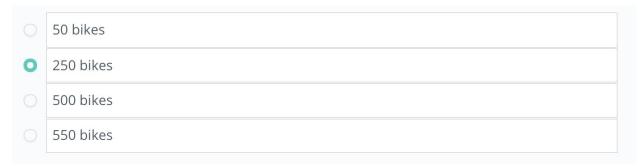
	Day of Week
0	Temperature
	Humidity

Rationale:

Temperature seems to have the strongest correlation with count, so it would probably do the best job predicting the number of bike rentals.

Question 4 - Draw Conclusions Step

Based on this graph of regressing bike rental count on temperature, how many additional bikes do you think would be checked out if the temperature rose from 2 degrees celsius to 30 degrees celsius?



Rationale:

Since the correlation isn't that strong, this prediction would probably be weak. However, the line of best fit is a good place to start if this scatterplot is all we have for our guess.

Question 5 - Communicate Step

What would be valid methods of communicating your conclusions from the Bike Sharing data?



Rationale:

Explaining the most important features to consider when predicting bike rental count would address our questions about this dataset, and a written report would be one way to communicate these results thoroughly.