a) What is open data? Given an example of open data that you produce which others can use? [2 + 2 = 4]

Open data is information that can be freely used and redistributed by anyone, no matter what they use it for. There are typically very few restrictions as long as it is legally open. An example of open data that we can produce is open source code, such as repositories and databases that you created to share.

b) You are analyzing a dataset and some attributes are missing.

b.1) What could be any 2 reasons why they are missing? [2 + 2 = 4]

One reason could be incorrect data entry. This would be a case of human error, simply due to misinputting the data. Another reason could be that the respondent chose not to answer, or they weren't available at the time. This would make it invalid since there would be no input.

b.2) What are any 2 ways you can still proceed with data analysis despite the missing values. For each, mention what assumption you are making and what are its risks. [(2+2+2) * 2 = 12]

List deletion - any rows that are missing values, you delete. We are assuming the data is missing at completely random patterns and wouldn't affect the observed data. The risk is that it could be biased and that the data sample size could drastically decrease, or it could change the data composition.