1. What is the goal of EDA (exploratory data analysis)?

Answer: The goal of EDA is to clean data, explore insights between relationships, and engineer features that would be helpful for the scope of the project that we are working on.

2. Suppose that you are given a dataset of customer product reviews for an e-commerce company. Each review is scored as a Likert-style survey item where 1 indicates negative sentiment about the product and a 5 is positive. These reviews are collected on the company's website. What problems do you expect to find in the raw data?

Answer: I would expect some neutrality bias. Depending on the distribution of the survey I would expect incorrect values to be entered, the data might also be in the wrong type. There may be duplicate entries, white space issues, and partially filled in rows.

3. If your task is to build features that give information about customer sentiments, how would you approach this task and what kind of methods would you apply to accomplish it?

Answer: My first task would be to define customer sentiment in the context of this task. Next would be to determine the features that would give me insight into this. Finally, I would apply a weight to each vector based on the amount of “sentiment” we could derive from that vector and create a vector using an average of “sentiment” gathered by each observation.

4. Try to identify some potentially useful features that you might derive from the raw data. How would you derive them and how would you assess the usefulness of those features?

Off the top of my head, I can see the weighted average of the most important features for the business as being useful. I would create this feature with the steps highlighted in my last answer.

Another interesting feature if the customer name is used is, we could tie a weight to how important they are as a customer to get a feel for the thoughts of who we think is important regarding us.